Transcriptomic investigation of the *Wolbachia* symbiosis in larval stages of *Brugia malayi*

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor of Philosophy

Shannon Quek

September 2019

Abstract

The *Wolbachia* genus of bacteria is comprised of obligate intracellular endosymbionts that are known to infect arthropods and nematodes. Most filarial nematodes of humans host maintain *Wolbachia* endosymbionts in a mutualistic association that is essential for nematode development, reproduction and the longevity of the adult parasites. As a result, much research has gone into investigating *Wolbachia*'s role in adult nematodes, both in understanding the basis of the mutualistic relationship, as well as exploiting the endosymbiont as a target for treatment. Less attention has been applied to understanding *Wolbachia*'s role in the biology of larval stages of filarial parasites.

To better characterise *Wolbachia*'s roles during these larval stages, RNA-sequencing technologies were employed to investigate the relationship between the parasitic filarial nematode *Brugia malayi*, and its *Wolbachia* endosymbiont during larval development and microfilarial transmission. This first involved the development of a manually curated, revised annotation of the *Wolbachia* genome using gene expression data, further corroborated by RT-qPCR and proteomics experiments. Second, the transcriptomes for both nematode and *Wolbachia* were then investigated across two major nematode developmental stages: the two weeks immediately after nematode infection into the mammalian host spanning the L3 to L4 developmental moult, and following *Wolbachia* depletion from *B. malayi* microfilariae during transmission to the mosquito vector.

The reannotation of the *Wolbachia* endosymbiont genome resulted in the identification of 21 new protein coding genes, 5 instances of non-model translational events, and 3 functional RNAs. Several newly identified genes were predicted to be unique to the *Wolbachia* genus, with a potential role in *Wolbachia*-nematode interactions. The transcriptome of developing L3 to L4 stages demonstrates *Wolbachia*'s ability to undergo coordinated control over its carbon metabolism to enable rapid population growth. The consistent upregulation of metabolic pathways, such as haem, nucleotide biosynthesis and Type IV secretion systems, complements the nematode host transcriptome, which was focused predominantly on its own growth and development, as well as regulating *Wolbachia* population during the L4 stage. *B. malayi* microfilariae depleted of *Wolbachia* have a significantly reduced ability to infect the mosquito vector, with transcriptome analysis of treated and untreated nematodes identifying targeted downregulation of chitinase and V-type ATPase transcripts in the treated

group. These targeted changes likely have a role in the nematode's ability to successfully penetrate the vector's midgut or achieve exsheathment.

Taken together, these observations illustrate a complex and dynamic relationship that *Wolbachia* has with its nematode host, expanding to more than just a mutualist important for adult parasite longevity and reproduction.

Table of Contents

Table of Figures

Table of Tables

Table of Abbreviations and Acronyms

ALB	Albendazole
ABC	ATP Binding Cassette
CD	CuffDiff
DAVID	Database for Annotation, Visualisation, and Integrated Discovery
DS	DESeq2
DEC	Diethylcarbamzine
ER	EdgeR
EST	Expressed Sequence Tags
FR3	Filariasis Research Reagent Resource
FAD	flavin adenine dinuleotide
L4	Fourth-stage larvae
FPKM	Fragments per Kiobase of exon per Million reads
GO	Gene Ontology
GSEA	Gene Set Enrichment Analysis
GAGE	Generally Applicable Gene Enrichment
GST	Glutathione S transferase
IGV	Integrated Genome Viewer
IVM	Ivermectin
KEGG	Kyoto Encyclopaedia of Genes and Genomes
MDA	Mass Drug Administration
Mbp	Megabase pairs
mf	Microfilariae
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
ORF	Open Reading Frame
PAL	Peptidoglycan-Associated Lipoproteins
PCR	Polymerase Chain Reaction
PC	Preventative Chemotherapy
GPLEF	Programme to Eliminate Lymphatic Filariasis
PRF	Programmed Ribosomal Frame-shifting
PDB	Protein Data Bank
PPDK	pyruvate phosphate dikinase
qPCR	Quantification Polymerase Chain Reaction
RT-qPCR	Reverse-Transcription Quantification Polymerase Chain Reaction
SMRT	Single Molecule, Real-Time
SNPs	single nucleotide polymorphisms
SNVs	single-nucleotide variants
SCR	Stop-Codon Read-through
L3	Third-stage larvae
TCA	Tricarboxylic Acid
VCF	Variant Call File
*w*AlbB	*Wolbachia* of *Aedes albopictus*
*w*Bm	*Wolbachia* of *Brugia malayi*
*w*Cle	*Wolbachia* of *Cimex lectularius*
*w*Pip.Mol	*Wolbachia* of *Culex pipens molestus*

*w*CqPip	*Wolbachia* of *Culex quinquefasciatus*
*w*DacA	*Wolbachia* of *Dactylopus coccus* (Supergroup A)
*w*DacB	*Wolbachia* of *Dactylopus coccus* (Supergroup B)
*w*Dcitri	*Wolbachia* of *Diaphorina citri*
*w*Dim	*Wolbachia* of *Dirofilaria immitis*
*w*Dana	*Wolbachia* of *Drosophila ananassae*
*w*Mel	*Wolbachia* of *Drosophila melanogaster*
*w*MelPop	*Wolbachia* of *Drosophila melanogaster* (strain popcorn)
*w*Ha	*Wolbachia* of *Drosophila simulans*
*w*No	*Wolbachia* of *Drosophila simulans* (strain Noumea)
*w*DSuzu	*Wolbachia* of *Drosophila suzukii*
*w*Fol	*Wolbachia* of *Folsoma candida*
*w*Gmm	*Wolbachia* of *Glossina morsitans morsitans*
*w*Ls	*Wolbachia* of *Litomosoides sigmondontis*
*w*Uni	*Wolbachia* of *Muscidufax uniraptor*
*w*VitA	*Wolbachia* of *Nasonia vitripennis* (Supergroup A)
*w*VitB	*Wolbachia* of *Nasonia vitripennis* (Supergroup B)
*w*Nfe	*Wolbachia* of *Nomada ferruginata*
*w*Nfla	*Wolbachia* of *Nomada flava*
*w*Nleu	*Wolbachia* of *Nomada leucophthalma*
*w*Npa	*Wolbachia* of *Nomada panzeri*
*w*Oo	*Wolbachia* of *Onchocerca ochengi*
*w*Ov	*Wolbachia* of *Onchocerca volvulus*
*w*BaOb	*Wolbachia* of *Operophtera brumata*
*w*Ppe	*Wolbachia* of *Pratylenchus penetrans*
*wW*b	*Wolbachia* of *Wuchereria bancrofti*

Acknowledgements

First and foremost, I would like to thank my primary and secondary PhD Supervisors Dr Simon Wagstaff and Professor Mark Taylor. For their enthusiasm, encouragement, endless patience with me, insightful discussions on the direction of this work, as well as the many doors that they have opened for me, I am extremely grateful.

Secondly, I would like to thank Dr Gareth Weedall for our discussions during various stages of my PhD, and helping ease me into the world of Bioinformatics and programming during those very early days. Similar thanks also goes to Dr James LaCourse, whose early enthusiasm to have me demonstrate on the MSc bioinformatics courses let me build up the confidence and drive to better my skills for more effective demonstrating.

I would also like to thank the friends I have made in members, past and present, of the Liverpool School of Parasitology's Parasitism (now Tropical Disease Biology) Department, and members of the Anti-Wolbachia Consortium laboratory, for helping me settle into the institution, new environment, and my early forays into working with live parasites. The same thanks go to my fellow PhD students in the Department, acting as the proverbial rock to lean on and discuss topics that concern *actual* biology in *living* organisms, as well as more mundane topics outside of the laboratory and science. Particular mention goes to Mr Andrew Stevens and Mr John Archer for their help in animal work and finding things that always seem to go missing in the laboratory for me. Additional thanks goes to Dr Kelly Johnston and Dr Louise Ford for their discussions on laboratory work and parasite biology. Further thanks go to Ms Mary Creegan, my ‘adoptive Scouse mum’ for her support, encouragement, and administrative assistance. Special mentions also go to Dr Anfal Yousef and Dr Piyaporn Jirawatcharadech, who have helped remind me multiple times that there’s more going on in life outside of the Laboratory, and what’s in front of the computer screen.

Finally, I would like to thank my mother, father, and sister for their endless support over the years in a variety of matters, despite the distance that separates us. Without it, I could never have reached this point.

Contributor statements

In accordance to the Postgraduate Research Code of Practice of the University of Liverpool, the individual contributions for each chapter of this thesis are as follows:

Chapter 1: Shannon Quek wrote this chapter. Dr Simon Wagstaff and Professor Mark Taylor supervised writing.

Chapter 2: Shannon Quek wrote this chapter and conducted bioinformatic analysis with some assistance from Dr Gareth Weedall. RNA-sequencing data used in this chapter was generated by Dr Christina Bronowski and the Liverpool Centre for Genomic Resources, who conducted RNA extractions and sequencing respectively. Dr Simon Wagstaff supervised bioinformatic analysis, and supervised writing alongside Professor Mark Taylor.

Chapter 3: Shannon Quek wrote this chapter and conducted bioinformatic analysis with some assistance from Dr Gareth Weedall. RNA-sequencing data used in this chapter was generated by Dr Christina Bronowski and the Liverpool Centre for Genomic Resources, who conducted RNA extractions and sequencing respectively. Dr Simon Wagstaff supervised bioinformatic analysis, and supervised writing alongside Professor Mark Taylor.

Chapter 4: Shannon Quek performed scientific experiments and bioinformatic analysis. Proteomics data used for analysis was generated by Dr Gemma Molyneux. Dr Simon Wagstaff supervised experimental design, work, and supervised writing alongside Professor Mark Taylor.

Chapter 5: Shannon Quek conducted RNA extraction and processing experiments, qPCR validation of targets, bioinformatic analysis, and wrote this chapter. Library preparation and sequencing was conducted by the Liverpool Centre for Genomic Resources. Animal maintenance, infection, and necropsy was conducted by Andrew Stevens, John Archer, and Amy Marriott. Mosquito transmission experiments were conducted by Dr Darren Cook. Dr Simon Wagstaff and Professor Mark Taylor supervised experimental design, work, and writing.

Chapter 6: Shannon Quek wrote this chapter. Dr Simon Wagstaff and Professor Mark Taylor supervised writing.

1. Introduction

1.1 Filarial diseases of humans

Parasitic filarial nematodes are the causative agent of several severely debilitating diseases in humans and animals, with a total of 8 different species of nematodes known to infect humans. These include *Wuchereria bancrofti*, *Brugia malayi*, *B. timori*, *Onchocerca volvulus*, *Loa loa*, *Mansonella perstans, M. ozzardi*, and *M. streptocerca*. To date, the primary focus of Global Neglected Tropical Disease programs has been on the control and more recently, the elimination of the first 4 named species of filarial nematodes[1,2], which together cause 2 major diseases in humans: lymphatic filariasis, and onchocerciasis. For these two diseases alone, recent estimates by the Global Burden of Disease study show a combined disease prevalence of just over 108 million people globally[3].

1.1.1 Lymphatic filariasis

Lymphatic filariasis (LF, also known colloquially as elephantiasis), is a parasitic disease caused by 3 different species of parasitic filarial nematodes; *W. bancrofti*, *B. malayi*, and *B. timori*. *W. bancrofti* is the predominant cause of LF infections throughout the world and can be found within tropical countries across South America, Central Africa, South-East Asia, and the Caribbean[1,4,5]. The distribution of *B. malayi* and *B. timori* by contrast is restricted to India and South-East Asia[4]. As of 2018, 51 countries across the World are classified as endemic for the disease[1,6], with a maximum estimated 71 million people suffering from the disease[7] (**Figure 1-1**). Infection with these nematodes often does not result in any overt pathology[4,5]. Instead, the localisation and migration of the nematodes within the infected individual results in significant inflammation and damage to the lymphatic systems, which occurs before overt symptoms appear[8]. The damaged lymphatic systems can progress to more clinically obvious symptoms that are directly debilitating, such as lymphoedema and hydrocele [4,8].

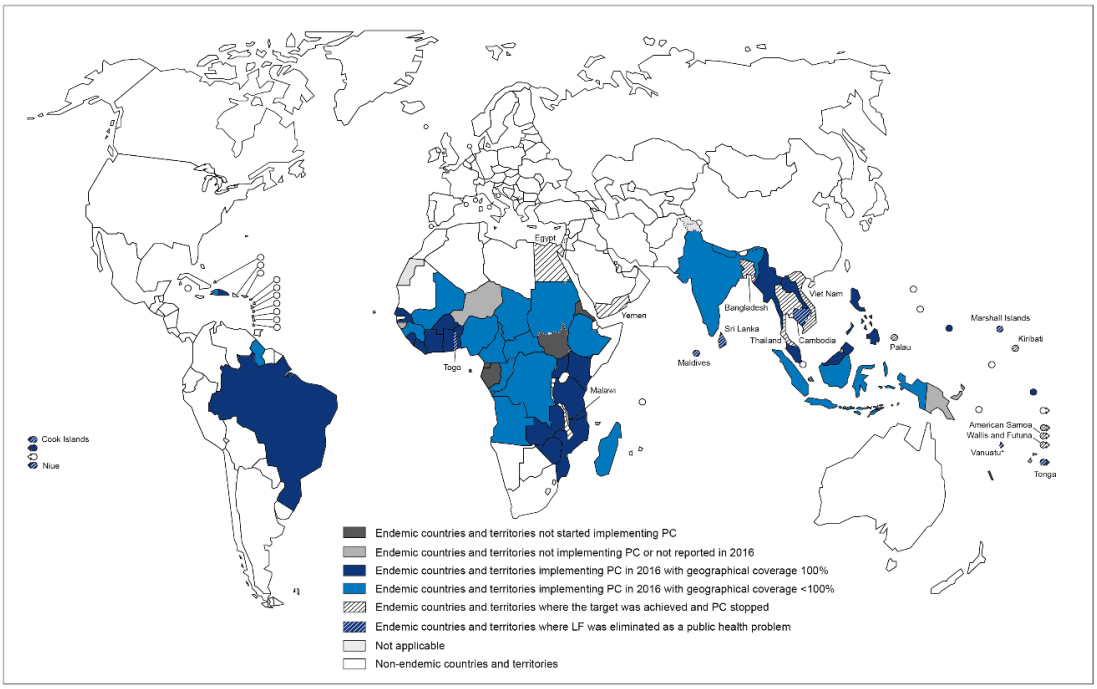


Figure 1-1: Distribution of LF across the world, overlaid with current status of preventative chemotherapy (PC), as of 2016. *Image adapted from the World Health Organization[9] on the 26th July 2019*

The life cycles of these 3 nematodes share much in common, with all beginning life as microfilariae (mf) encased within a chitinous sheath, circulating within the blood of an infected host. These can then be transmitted to a variety of mosquito vector species, including members of the *Aedes*, *Anopheles*, *Culex*, and *Mansonia* genera[5], although vector competency can vary significantly by strain[5,10,11]. Subsequent to uptake via a blood meal, mf then penetrate the midgut of susceptible mosquito vectors, shedding their sheaths in the process[10–12], and migrate through the haemocoel to the flight muscle cells of the vector[13]. Here the nematodes develop over a period of approximately 2 weeks[5], and progress through two developmental moults to become known as third-stage (L3) infective larvae[5], which migrate to the mouthparts of the mosquito. Upon the next blood meal, the L3 escape from the mosquito mouthparts, and enter the wound caused by the vector. The L3 then migrate into the lymphatics system, before undergoing two additional moults: the first within 2 weeks of infection, and the second after 2 months, with reproductive maturity and release of new mf taking up to a year after initial infection[14,15]. Nematodes remain reproductively active for between 5-8 years[4], during which they produce thousands of microfilariae a day that migrate to the blood system and complete the life-cycle (**Figure 1-2**).

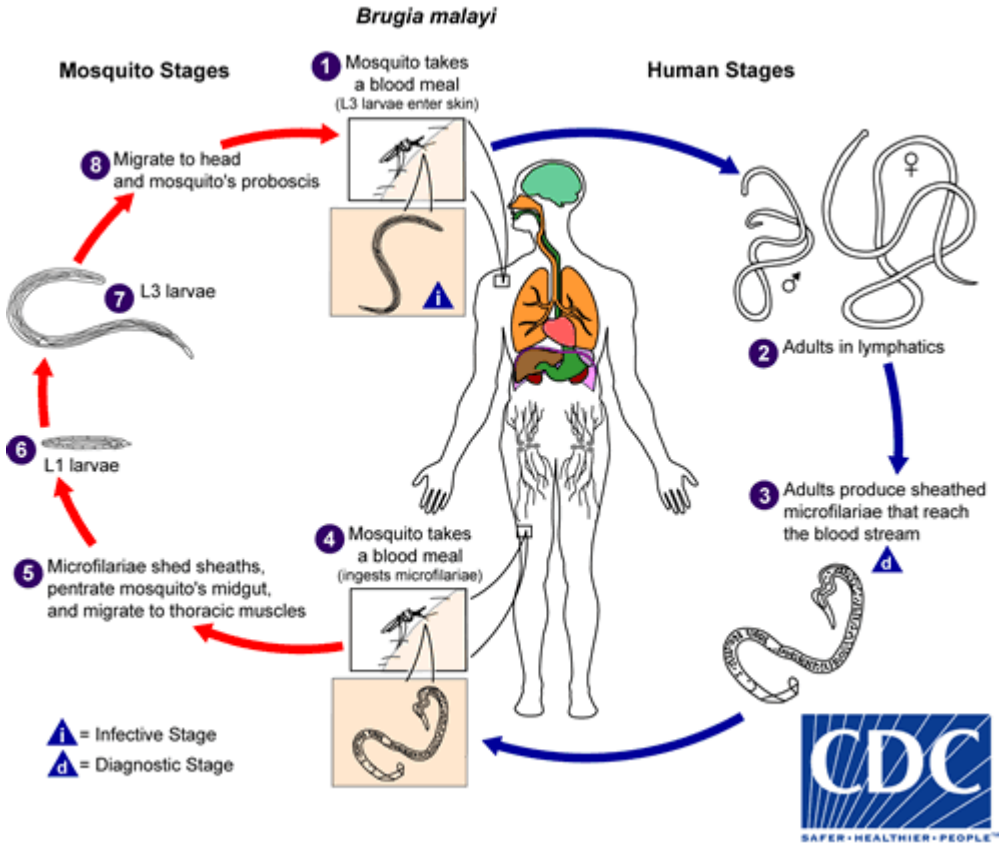


***Figure 1-2: Depiction of the life cycle of* Brugia malayi.** *Taken from the Centers for Disease Control and Prevention[16], accessed on 29th July 2019.*

1.1.2 Onchocerciasis

Human onchocerciasis is caused by a single nematode species, *O. volvulus*. A total of 31 countries across the World are classified as endemic for this disease, with the majority of them focused across much of Sub-Saharan Africa, and some parts of South America and Yemen[2] (**Figure 1-3**). Clinical manifestations of onchocerciasis are due primarily to host immune responses towards dead or dying microfilariae[4,17]. As mf of these species typically localise and migrate through subcutaneous tissues, this frequently results in inflammatory-mediated skin diseases that range from troublesome itching and acute, chronic papular dermatitis, to depigmentation of skin, as well as a loss of skin elasticity and structure[4,17]. These mf are also known to migrate through ocular tissues, with inflammation to dead mf resulting in visual impairment, and eventual blindness of patients[17]. There have also been studies which indicate the presence of different strains of *O. volvulus* separated by geographical area, with areas of Africa predominantly occupied by savannah regions associated more with blindness than areas predominantly occupied by forested regions, which are more associated with skin disease[17,18].

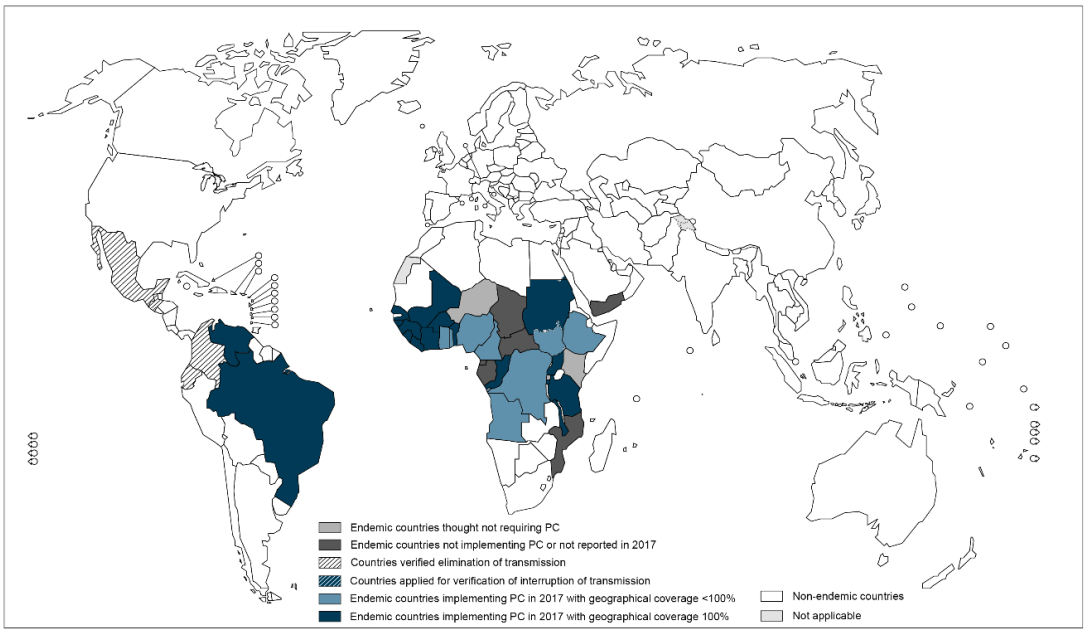


Figure 1-3: Distribution of* Onchocerca volvulus *across the world, overlaid with current status of PC, as of 2016.
Image adapted from the World Health Organization[19] *on the 26th July 2019*

Unlike the mf of LF-causative nematodes, mf of *O. volvulus* do not maintain a chitinous sheath. Their cutaneous localisation allows for transmission to the blackfly vector of the genus *Simulium*, which take up mf during a blood meal. Mf then penetrate through the blackfly midgut, before migrating through the haemocoel and into the thoracic muscles. The mf then undergo 2 moults to become infective L3-stage larvae over the next week[17], similar to filarial nematodes that cause LF. These L3-stage larvae then migrate to the mouthparts of the blackfly, where upon taking another blood meal, the L3 larvae escape from the blackfly and migrate into the bite wound. The L3's then undergo two additional moults- one within a week after initial infection, the second between 1 to 3 months after infection, before reaching reproductive maturity[17]. These adults reside in fibrous, highly vascularised nodules in subcutaneous and deeper tissues, releasing ~1,500 mf a day for up to 11 years[17]. These mf then migrate through the subcutaneous tissues, repeating the cycle again (**Figure 1-4**).

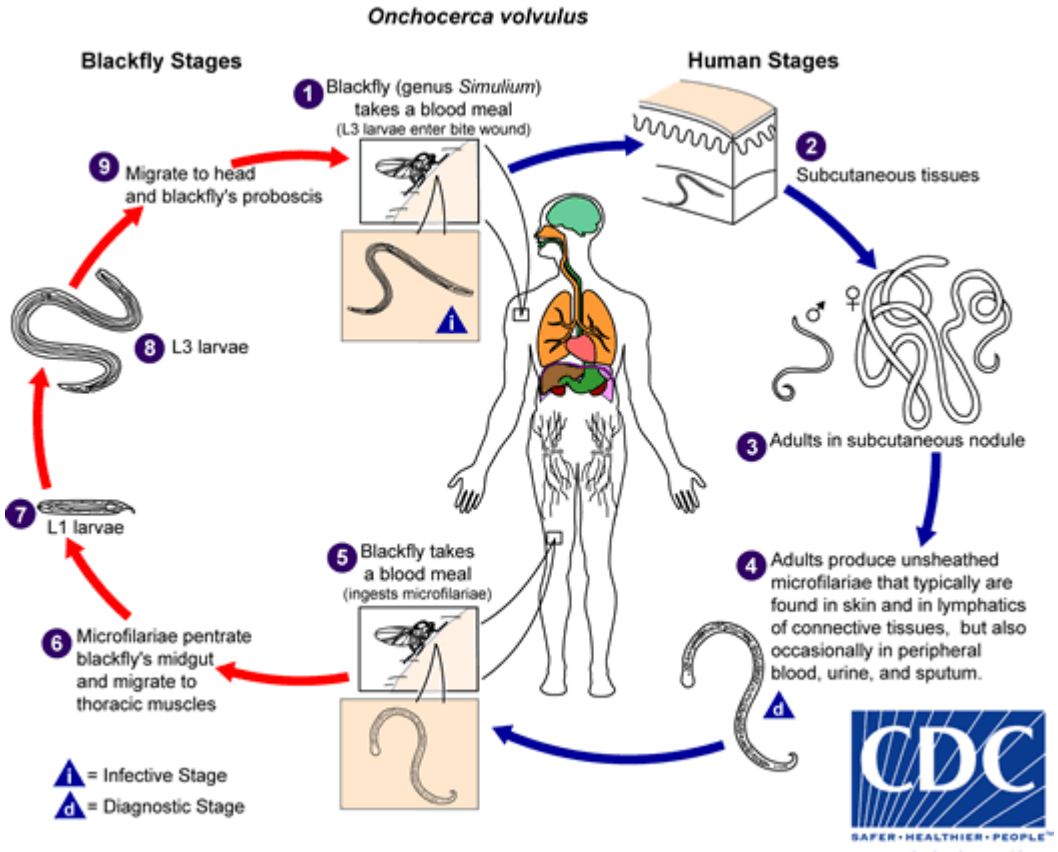


***Figure 1-4: Life cycle of* Onchocerca volvulus.** *Taken from the Centers for Disease Control and Prevention[20], accessed 29th July 2019*

1.1.3 Current treatment strategies

Treatment for these filarial nematodes as advocated by the World Health Organisation (WHO) involves mass drug administration (MDA) of a combination of ivermectin (IVM), albendazole (ALB) and/or diethylcarbamazine (DEC) at least once yearly for a minimum of 5 years in endemic regions for lymphatic filariasis, or a minimum of 15 years of IVM for regions endemic for onchocerciasis[1,2,4,21]. The drug combination varies depending on the geographical distribution and co-endemicity of other filarial species, and primarily target and kill the mf of the nematodes (microfilaricidal activity), thus blocking transmission. The Global Programme to Eliminate Lymphatic Filariasis (GPELF) was launched in 2000 with an initial goal of global elimination by 2020[1,21]. However, the present coverage of MDA is estimated at only 40%, with only a resultant 30% decline of disability adjusted life years since 1990 highlights the extent of disease burden still present[22]. One explanation for the continued persistence of LF is that current drugs do not target the adult worms (macrofilaricidal activity) which continue to survive, breed and reproduce[23]. Therefore, MDA regimens that incorporate these drugs are designed to suppress mf populations and prevent transmission, whilst allowing mature adult nematodes to die via their natural life cycle[23].

This strategy does pose certain problems, not least of which is the risk of resistance arising to these drugs[24,25], with the earliest reports from 2004 by Awadzi *et al.* observing certain adult female *O. volvulus* nematodes that did not respond adequately to multiple doses of Ivermectin[26]. In addition, IVM and DEC are known to cause severe adverse effects in individuals co-infected with *L. loa* due to the rapid killing of mf of these species, and subsequent sequestration and blockage of micro-vasculature, leading to encephalopathy[27]. This increases the logistical, screening and monitoring requirements before the appropriate treatment regimen can be selected. Furthermore, experience of adverse events are strongly associated with non-compliance to IVM treatment[28]. Simultaneously, external geopolitical factors that may arise during the course of MDA can easily disrupt infrastructure or transport logistics that must be put in place to perform these annual MDAs. Alternative treatment strategies or targets are thus of significant importance if the disease is to be eliminated within acceptable timeframes. One such alternative target is an intracellular bacterium with a mutualistic symbiotic association with these filarial species, known as *Wolbachia pipientis*.

1.2 Biology of the obligate intracellular bacterium *Wolbachia*

Wolbachia pipientis is an obligate intracellular, gram-negative alpha-proteobacteria of the order Rickettsiales, transmitted vertically via the maternal germline[29,30], and usually found within host-derived vacuolar membranes[29,30]. It is the sole member of the genus *Wolbachia*, and notable for infecting a large number of different arthropod species, as well as a selection of filarial nematodes of both medical, veterinary, and to a limited extent agricultural, importance[23,31–33]. With such a broad range of host species, comes an equally broad range of effects that the *Wolbachia* endosymbiont elicits within their hosts. Such effects include parthenogenesis (asexual reproduction), feminization of genetic males, or cytoplasmic incompatibility (CI), where intraspecies crosses of arthropods infected with different strains of *Wolbachia*, or crosses of *Wolbachia*-infected male arthropods with uninfected females, fail to result in viable offspring[31,34]. As of 2019, these bacteria are classified as members of different 'supergroups' within the genus, designated A through to Q[35,36]. An exception to this is supergroup G, which has been decommissioned[36] due to concerns the genes used to classify the supergroup were the result of recombination between supergroups A and B, and may not be a unique clade in its own right[37]. Classification between supergroups was traditionally based upon phylogenetic analyses of one or more *Wolbachia* genes[30,35,38,39]. This included, but was not limited to, 16S rDNA, *ftsZ* (cell division protein), *wsp* (*Wolbachia* surface protein), or *gltA* (citrate synthase). Developments in high-throughput genetic sequencing has allowed for significantly higher resolution of the *Wolbachia* species

boundary, and has led to recent efforts to more accurately categorise the supergroups into distinct species, with some pre-existing supergroups possibly having multiple species within them[40,41]. Whilst these new categorisations have been proposed, none have yet to be formally adopted.

1.2.1 Distribution, history, and role of *Wolbachia* amongst arthropods

Wolbachia was first discovered in 1924 when Hertig and Wolbach identified the presence of intracellular bacteria within the reproductive tissues of the mosquito *Culex pipiens*[42], which were eventually named *Wolbachia pipientis*[43]. *Wolbachia*'s presence has since been identified in a wide range of arthropods[31] and has been predicted to be present in the majority of all arthropod species, with Hilgenboecker *et al.*[44] predicting that over 65% of arthropod species play host to *Wolbachia*. These estimates make the *Wolbacha* genus one of the most abundant intracellular genera of bacteria currently known[32]. Interestingly, although *Wolbachia* has been predicted to infect such a broad range of arthropod species, the effective *Wolbachia* infection rate within species varies between the extremes. Various studies have predicted a 'most-or-few' pattern, where either individuals within an arthropod species have either a near complete infection rate, or are barely infected[44,45].

The association between *Wolbachia* and their reproductive parasitism of arthropods, most notably in the form of cytoplasmic incompatibility, has since become the hallmark of *Wolbachia*'s presence. These induced changes are predicted to aid *Wolbachia*'s spread through host populations due to their maternal inheritance pattern[31]. Despite this, there has been a growing body of work that suggests *Wolbachia* within arthropods may play an additional, more facultative role outside of simply being a modulator of reproductive processes (as reviewed by Zug and Hammerstein[34]). Such roles have been observed to include nutritional mutualism, particularly in iron acquisition to benefit host fecundity[46,47]. Additional roles include increasing resistance against infection of arthropod-pathogenic viruses[48–51], or human viruses that use the arthropod as a vector[52,53]. In an extreme case, *Wolbachia* of *Cimex lectularius*, or the common bedbug (*w*Cle, member of supergroup F), has become an obligate endosymbiont for arthropod fecundity, development and survival[54,55]. In this relationship, *Wolbachia* has been shown to act as a nutritional mutualist that provides vitamin B to allow the bedbug to successfully develop through its instar moults, as well as reproduce successfully[54,55].

1.2.2 Distribution and history of *Wolbachia* amongst nematodes

The presence of bacteria infecting the hypodermal tissues of filarial nematodes was first identified in 1975 by McLaren and Worms[56] in *Dirofilaria immitis* (the causative agent of dog

heartworm) and *Brugia pahangi* filarial nematodes (a close relative of *B. malayi* that infects dogs and cats). Despite the identification of the presence of these bacteria, these observations were left mostly ignored until much closer to the beginning of the 21st century, when Sironi *et al.* in 1995 were able to successfully identify the bacteria in *D. immitis* as members of the *Wolbachia* genus by sequencing of its 16S rDNA[57]. Subsequent studies identified *Wolbachia*'s presence in tissues of other filarial nematodes, many of which are parasites of humans[30]. Examples of these include members of the *Onchocerca*[29], *Brugia*[58,59], and *Mansonella* genera[60,61]. Until this point, *Wolbachia* was believed to only be present in arthropod species.

Wolbachia's distribution appears limited to the subfamilies Onchocercinae and Dirofilariinae[35,38,62,63], although more recent studies have identified a *Wolbachia* presence within the Pratylenchidae family of plant parasitic nematodes, such as *Pratylenchus penetrans* or *Radopholus similis*[33,64,65]. Unlike *Wolbachia* of arthropods, which are distributed over >10 supergroups, *Wolbachia* of nematodes have been separated to just 5 different supergroups: Supergroups C, D, J (which are comprised solely of *Wolbachia* of medically and veterinary-important parasitic nematodes, supergroup F, (which contains *Wolbachia* of both nematodes and arthropods that exhibit a predominantly commensal phenotype, and supergroup L (which contains *Wolbachia* of plant parasitic nematodes of the Pratylenchidae family[65,66]). In addition to this significantly narrowed range of hosts, *Wolbachia* of filarial nematodes, in species that harbour them, have been noted to be present in 100% of nematode individuals studied, suggesting that their maintenance within the nematode is obligate, as reviewed by Taylor *et al.*[67]. The only possible exception to this is *Mansonella perstans*, with different studies both confirming[60,68] and denying[69] the presence of *Wolbachia*.

Perhaps because of this apparently obligate requirement of *Wolbachia*'s presence in nematodes, questions have arisen as to why *Wolbachia* is not present in all members of the Onchocercidae, such as *L. loa*, or other *Onchocerca* species that parasitize animals. Several hypotheses that may explain this phenomenon involve the possibility of separate acquisitions in different nematode species over the course of evolution, or that *Wolbachia* infection occurred by a single ancestor, before being lost over time[23].

1.2.3 Population dynamics and localisation during filarial nematode life cycles

McGarry *et al.* in 2004[15] were able to identify the quantity and localisation of *Wolbachia* in the nematode *B. malayi*'s lifecycle, with time-points studied including the mf, L2 and L3 stages, as well as tracking L3 development through the L4 stage, and into reproductive

adults[15] (summarised in **Figure 1-5**). *Wolbachia* populations were quantified via real-time quantitative Polymerase Chain Reaction (qPCR) techniques. This used a normalised ratio between the *Wolbachia* surface protein gene (*wsp*) and *B. malayi* glutathione S-transferase gene (*gst*), both of which the study had determined existed as single-copies within the organisms[15]. This technique highlighted that *Wolbachia* populations are typically present at a low level (approximately 100-200 bacteria per nematode) during the early life cycle stages (mf through to L3).

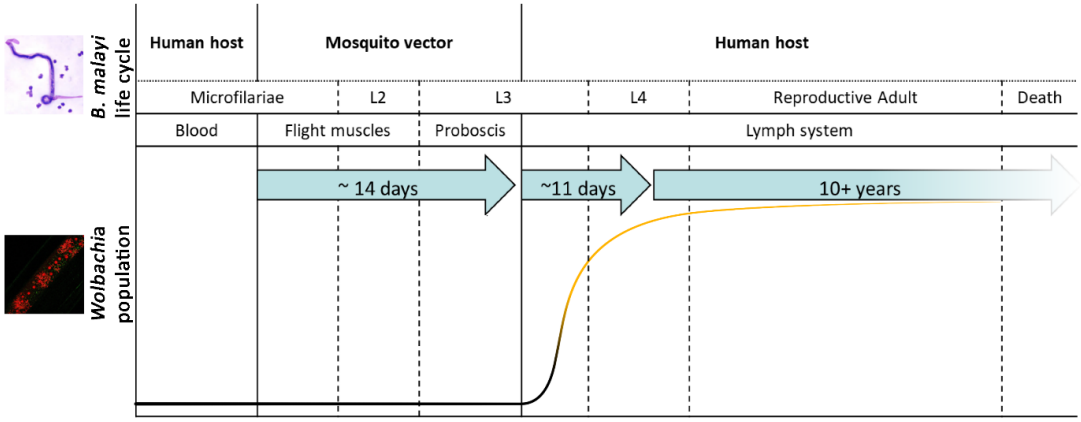


Figure 1-5: Schematic overview of the* Wolbachia *population across the life cycle of a* B. malayi *nematode host, not drawn to scale. *Note the very low population levels during the first few life cycle stages, before infection into the final mammalian host. Also note that in a matter of days,* Wolbachia *populations expand rapidly, before the nematode reaches the L4 developmental stages. Images of* Wolbachia *taken from Taylor et al.*[67]*, and* B. malayi *taken from the Centers for Disease Control and Prevention*[70]

Within the first week after infection of the mammalian host, *Wolbachia* populations were noted to increase by approximately 600-fold[15]. This population growth was noted to continue throughout the nematode's lifespan within females, likely owing to the continuous growth in size of the nematode as well as accumulation of new embryos and mf, and their own *Wolbachia* populations, within the nematode reproductive tract[15]. By contrast, adult male nematodes appeared to maintain a consistent *Wolbachia* population throughout the remainder of their lifespan[15].

Electron microscopy found these *Wolbachia* to be localised within the developing hypodermal chords of the L3 developmental stage onwards[15]. This continues until 21 days after infection (by now at the L4 stage) where the hypodermal chords are properly developed, with none found within the gonad tissue. In older, reproductively active female nematodes, *Wolbachia* could still be observed within the hypodermal chords in large numbers, as well as the ovaries and intra-uterine developmental stages. No such invasion of *Wolbachia* could be seen in male testes or sperm[15]. These observations in terms of *Wolbachia* population changes and/or their localisation, was further corroborated in a

subsequent study by Fenn and Blaxter in 2004 using similar techniques[71], and by Fischer *et al.* in 2011, focusing on microscopy techniques[72]. The latter publication was also able to identify *Wolbachia* infecting the testis of male nematodes , but never within the spermatozoa or mature spermatids[72].

Within adult females, germline invasion of *B. malayi* nematodes by *Wolbachia* was observed and tracked by Landmann *et al*.[73] This showed how the first zygote cleavage parallels that in the model nematode *Caenorhabditis elegans* to generate the somatic AB precursor cell, and the smaller P1 germline precursor cell, with *Wolbachia* preferentially localizing in the P1 blastomere, rather than the AB blastomere. *Wolbachia* localised within the P1 blastomere would eventually become concentrated in germline cells, as well as cells that form the hypodermis of the microfilariae, the latter of which will eventually develop into the hypodermal chords.

1.2.4 Role of *Wolbachia* in their nematode hosts- pathology and survival

Since this initial discovery, it is now recognised that the majority of known parasitic filarial nematodes that infect humans maintain a *Wolbachia* endosymbiont, with the exception being *L. loa*[74]. In contrast to the varied phenotypes these *Wolbachia* bacteria cause in arthropods, the presence of *Wolbachia* in nematodes has been shown to be obligate, and required for successful nematode growth, fertility, and long-life span. This has been shown via a variety of studies which looked at the effects of antibiotic treatment, and subsequent *Wolbachia* depletion, on filarial nematodes.

Antibiotic treatment of filarial nematodes was first shown to have inhibitory effects on both the development of nematodes, as well as a halt in embryogenesis, by Bosshardt *et al.*[75] in 1993. The association of this phenotype to the depletion of the *Wolbachia* endosymbiont was not fully established until 1999. Here, Hoerauf *et al.*[76] utilised mice and Mongolian jird animal models infected with *Litomosoides sigmodontis*, a parasitic filarial nematode of rodents that harbours *Wolbachia*, and treatment with tetracyclines for up to 41 days immediately after infection. The authors subsequently noted a block in nematode embryogenesis, a minimum of 4-months of nematode infertility, as well as stunted growth and development. Treatment with other classes of drugs that are not known to affect *Rickettsiales* bacteria did not show this effect. The authors also showed that *Acanthocheilonema vitae*, another filarial parasite that does not maintain a *Wolbachia* population, did not suffer any deleterious effects upon treatment with tetracycline antibiotics[76]. A similar study by Bandi *et al.*[77] in 1999 also confirmed similar observations in the filarial nematodes *Dirofilaria immitis* and *B. pahangi*. Using cattle infected with *O.*

ochengi, Langworthy *et al.*[78] in 2000 successfully identified a reduction in adult male and female nematode viability, via reduction in their motility, as well as a reduction in size, and clearance of, nodules after treatment with oxytetracycline. Furthermore, they also identified a significant reduction in fecundity of adult females following treatment, with an observable reduction in developing embryo quantities, as well as an increased rate of malformed embryos[78]. These described studies provide some of the first evidence of macrofilaricidal activity via depletion of the *Wolbachia* endosymbiont.

Experiments to confirm *Wolbachia* as a valid target in filarial nematodes of humans were first done by Hoerauf *et al.*[79] in 2000, using doxycycline treatment of *O. volvulus* nematodes. These results again reflected that of early studies, namely that a relatively short (6-week course) of antibiotics treatment, in this case doxycycline, was sufficient to achieve long-term sterility of adult female nematodes, as well as degeneration, or shortened lifespan of adults in general. This was followed by a series of additional studies in the filarial nematode species *B. malayi*[80] and *W. bancrofti*[81,82] that showed similar results.

There has also been a limited amount of work done investigating the effects of *Wolbachia* depletion within the microfilariae stages- a point in the nematode life cycle where the *Wolbachia* population is comparatively small (**Figure 1-5**). Initial studies by Sucharit *et al.*[83] as early as 1978 first identified that tetracycline treatment of *B. pahangi* mf, before infection in *Aedes togoi* resulted in a reduced recovery rate of the infective L3 stage. At the time, *Wolbachia*'s presence in nematodes was yet to be established. Subsequent experiments by Srivastava and Bhattcharya[84] in 2003, Arumugam *et al.*[85] in 2008, and Albers *et al.*[86] in 2012 utilising either tetracycline or doxycycline treatments further corroborated the observation of a reduction in mf that successfully developed to the L3 stages in *L. sigmodontis, B. malayi*, and *O. volvulus* respectively. Only the latter 2 studies were able to directly attribute reductions in L3 recovery rate to *Wolbachia* depletion however, and the exact mechanisms behind this has yet to be elucidated.

Outside of these direct effects on the nematode, *Wolbachia* has been implicated in actively contributing to pathology seen in patients, primarily via the induction of strong immune and inflammatory responses against various *Wolbachia* proteins, such as *wsp* and peptidoglycan-associated lipoproteins (PAL)[87–90]. Injection of *Wolbachia* of arthropods, or *Wolbachia*-containing extracts from *B. malayi* or *O. volvulus*, has been shown to elicit the recruitment of neutrophils in a mouse model of ocular onchocerciasis. Conversely, injections of nematode extracts depleted of *Wolbachia*, or extracts of *A. viteae* (which does not have *Wolbachia*)

showed significantly milder responses[87,91,92]. This recruitment results in a feedback loop, with the constant recruitment of additional neutrophils, and the eventual disruption of normal corneal clarity[91,92]. This strong immune reaction elicited by *Wolbachia* has also led to the generation of hypotheses that the intracellular endosymbiont may act as a defensive mutualist, aiding in the evasion of the host immune system. For instance, studies in *Onchocerca* spp. nematodes results in the recruitment of neutrophils- white blood cells that are usually employed for an antibacterial immune response, to infected regions, or regions injected with filarial extracts containing *Wolbachia*[93,94]. These neutrophils were shown to be replaced by eosinophils after *Wolbachia* depletion by antibiotics, permitting a more effective immune response against the host nematode[94], possibly resulting in the macrofilaricidal activity. A similar effect has been observed in *B. malayi*, where the strong immune response of the host against *Wolbachia*, coupled with long-term exposure, eventually results in a significantly dampened immune response[95].

1.2.5 *Wolbachia* as an alternative treatment target for human parasitic filarial nematodes

These studies into the essentiality of *Wolbachia* for their nematode host's survival, as well as their contribution to pathology, represented a significant advance in treatment options, when compared to standard anthelminthic drugs IVM, ALB, and/or DEC. Antibiotic regimens using doxycycline were tested in clinical trials, and indicated that up to 8-week long treatment courses were sufficient to induce macrofilaricidal effects and a block in embryogenesis[96]. A series of regimen reduction studies were conducted in an attempt to reduce the time required for treatment, as well as examine whether a combination of different drugs, including antibiotics, could achieve better results[67,97–99]. Turner *et al.*[100] for instance showed that a 3-week course of treatment with a combination of doxycycline and ALB or IVM was sufficient to induce long-term amicrofilaremia in treated individuals (minimum of 24 months amicrofilaremia) through blockage of embryogenesis, but not enough to induce macrofilaricidal effects. Already, this short course of treatment shows improvements than standard treatment with ALB or IVM. Supali *et al.*[101] 2 years later showed that 6-weeks of treatment with a combination of doxycycline and ALB or DEC resulted in sustained reduction of microfilariae, macrofilaricidal effects, as well as a reduction in adverse events experienced by the patient.

Optimising the treatment regimen is only one stage of the process however, as tetracyclines are contraindicated in pregnant women and children aged 8 or less[102], thus necessitating the search for alternative drugs. The most recent efforts to identify and modify existing drugs, or

design entirely new drugs for the treatment of filarial nematodes via targeting *Wolbachia* have been spearheaded by the Anti-*Wolbachia* (AWOL) Consortium. Established in 2007 via funding from the Bill and Melinda Gates foundation, the consortium has identified a number of potential candidates that can achieve potent anti-*Wolbachia* activity within 7 days of treatment, or less[102]. One example is the antibiotic rifampicin, used to treat bacterial infections of *Mycobacterium tuberculosis* or *Legionella pneumophila*. Although early studies have shown that treatment with a standard regimen in humans (10mg/kg) was insufficient to achieve equivalent anti-*Wolbachia* activity to doxycycline[103,104], subsequent studies and pharmacokinetic modelling showed that a larger dose (30 to 40 mg/kg) would be sufficient, and safe in humans[105]. Turner *et al.* was ultimately able to show that 7 days of treatment with this enhanced dose of rifampicin plus ABZ was sufficient to induce both micro- and macrofilaricidal effects in pre-clinical models[106]. This is but one out of several thousand candidates that the AWOL consortium has identified in its decade-long history however[107,108]. In addition, there are 2 new anti-*Wolbachia* compounds that are being developed, or have entered clinical trials, that promise both specificity against *Wolbachia* as well as superior efficacy to currently known compounds: TylaMac (a modified macrolide) and AWZ1066 (a thienopyrimidine/qunazoline derivative) [109,110].

1.3 Application of Omics technologies to understand nematode-*Wolbachia* symbiosis

Whilst the exploitation of *Wolbachia* as a target for treatment of filarial nematode infection has been identified and refined since 1998, the basic biology underpinning the symbiotic relationship between the two organisms is still poorly understood. The advent of sequencing technologies allowed a more in-depth understanding of the role *Wolbachia* plays in the symbiotic relationship with their nematode host, starting with the genome sequences of both *Wolbachia*[111] and nematode host[112]. This review will first cover the developments in sequencing technologies, before reviewing how these developments have contributed to advances made in elucidating the relationship between *Wolbachia* and its nematode hosts.

1.3.1 History of sequencing technologies

As reviewed by Shendure *et al.*[113], the first major breakthrough in large-scale sequencing technologies was made by Sanger, Nicklen and Coulson in 1977[114]. The technique utilised the inability for di-deoxynucleotides to be incorporated into DNA sequences in the place of deoxynucleotides which are normally incorporated into DNA sequences. By substituting a small amount of regular deoxynucleotides for di-deoxynucleotides within DNA polymerization reactions, it was possible to generate DNA fragments of varying lengths based

on where the di-deoxynucleotide was incorporated, with the variable fragments run on a gel, and the underlying DNA sequence elucidated. This technique served to underpin early genome sequencing projects, most notably that of the Human Genome in 2001[115]. The next major breakthrough in sequencing technologies was not until the late 1990's to early 2000's, with the development of massively parallel, or Next Generation Sequencing (NGS) techniques.

Multiple companies have designed workflows and machines to exploit massively parallel sequencing techniques, as reviewed by Slatko *et al.*[116], but all are based on a 'synthesis by sequencing' approach. In essence, the DNA sequence of interest is first fragmented into small (typically less than 200 base-pair length) sequences, immobilized onto a medium, before *in-vitro* amplification to generate millions of target DNA templates separated into distinct clusters of the same template. The exact DNA sequence of these template clusters could then be determined via sequential washing of labelled nucleotides over the medium, with nucleotides releasing a detectable fluorescent signal upon incorporation into the immobilized clusters by DNA polymerases. An alternative technique that has recently gained prominence is Single Molecule, Real-Time (SMRT) sequencing by Pacific Biosciences, which rather than sequencing amplified small fragments of the target, allows entire lengths of large DNA sequences to be read directly instead. This utilises a gap less than half the wavelength of light in which an engineered DNA polymerase enzyme can be anchored within. As a target DNA template is allowed through, fluorescently labelled nucleotides are incorporated and the fluorescent signal then measured from the gap. This technique allows for direct sequencing of DNA strands up to, or exceeding, 50 kilobase-pairs long[113,116].

1.3.2 gDNA sequencing, assembly, annotation

One of the primary goals of most DNA sequencing projects is the eventual sequencing of an organism's genome. Knowledge of the genome allows for analysis of the various biochemical and signalling processes that the organism is able to undergo, by the identification of coding genes (either for proteins or functional RNAs) within the genome. DNA sequencing is only the first step towards this goal, as subsequent assembly of the reads into contigs and scaffolds are required based on the overlapping 'ends' of the sequences. Assembly of data generated by NGS techniques can be computationally very difficult due to the prevalence of repetitive, or duplicated, sequences within genomes of many organisms[117]. The short-read lengths mean that repetitive regions may be misidentified and result in incorrectly assembled, or even collapsed, genome sequences that omit fragments of coding regions[117]. These issues pose a challenge when studying the genome or transcriptomes of pathogenic

organisms that have global distribution, as such sequencing or assembly errors may not be easily separated from the backdrop of natural biological variability in the genomes of field-derived samples. Alternative techniques do exist to reduce these problems however, with techniques such as SMRT sequencing leveraging its long-read lengths to effectively sequence through repetitive regions of a genome, thus aiding genome assembly. Whilst the throughput and accuracy of SMRT technology is not as high as NGS techniques, sequencing projects that utilise combinations of the two techniques have generated high-quality genome assemblies for analysis[118,119].

Full exploitation of genome assemblies requires accurate annotation of genes and their function. This has typically been accomplished computationally thanks to algorithms developed for the large amount of sequence data that can be generated from a single project. Early programs that were designed for identifying such open reading frames relied on pre-set algorithms trained on model organisms. As time passed and more sequence data became available, such programs have become increasingly complex, incorporating information from multiple databases, as well as additional sequencing data in the form of RNA, to influence calls for coding genes[120]. However, new algorithms are often never reapplied to genomes that have been sequenced and annotated in the past. In a 'transfer of annotation' strategy, computationally assigned annotations for newly sequenced genomes are often inferred from homology, or the presence of conserved motifs or domains within the translated amino acid sequence, from databases of existing genes via the use of search engines such as BLAST or InterProScan[121,122]. If annotations from the source genome do not incorporate the latest knowledge, this computational assignment will only serve to propagate any errors, unless human curation of the results is actively involved, or new annotations are retroactively applied to source genomes.

1.3.3 RNA sequencing, assembly, and differential expression analysis

RNA sequencing relies on the same basic principles of DNA sequencing technologies, except utilising reverse-transcribed, complimentary DNA sequences compared to genomic DNA sequences. It is an alternative and complementary strategy to Microarray technology, the latter of which has predominantly been used to measure differential gene expression within a sample population, tissue, or organism, of interest (as reviewed by Bumgarner[123]). Both techniques allow for the measure of gene expression during a variety of experimental conditions. Examples could include environmental changes in temperature or chemical exposures, or internal life cycle processes, such as different phases of growth. Microarray technology however requires at least some knowledge of the organism of interest's genome

in advance for the designing of Expressed Sequence Tags (ESTs) for each gene in the genome. By contrast, RNA-sequencing does not require prior knowledge of an organism's genome - it could instead, be used to directly assemble a 'transcriptome' profile of the coding genes within an organism without one[124]. This lack of reliance on an existing genome sequence allows RNA-sequencing to be used for more than just analysis of differential expression, such as identifying the presence of previously unrecognised genes within a genome[123], or detection of novel splice variants of genes[123,125].

Specifically for analysis of differential expression between samples, sequenced RNA reads are aligned computationally, either against an assembled transcriptome, or against an existing genome. This alignment step would need to contend with sequencing errors, repetitive regions, nucleotide variants within either the genome or transcriptome, as well as splicing within eukaryotes[124,126]. Once aligned, gene expression and quantification can be estimated computationally, after being normalised for factors such as sequencing depth, the length of a particular gene, or controlled for the presence of gene isoforms[124,126]. After these steps, differential expression analysis can then begin via using tools such as EdgeR, DESeq2, or CuffDiff[127–130]. In general, differential expression analysis tools attempt to identify genes that show statistically significant differences in expression between given experimental conditions. These tools employ a number of statistical models and assumptions, primarily in an attempt to account for biological variability in samples[124,126]. EdgeR and DESeq2 for instance utilise variants of Poisson and negative binomial distributions respectively[127,128], and operate on the null hypothesis that there are no differentially expressed genes within a given dataset. After the identification of genes that show statistically significant differential expression, further biological insight can be gained via investigation of these candidates in the context of gene 'sets'. These gene sets are groups of genes that share a particular property, such as being members of a metabolic or signalling pathway, being associated with a specific Gene Ontology (GO) term[131,132], or showing co-expression during a given phenotype or biological event[133–135]. Determining if such gene sets may have more differentially expressed genes than what one would expect by chance, allows for a greater understanding of the underlying biology within the experimental samples.

1.3.4 Observations from the genome sequences of *Wolbachia* and filarial nematodes

The first genome sequencing projects of *Wolbachia* were performed by Wu *et al.*[136] in 2004 and Foster *et al.*[111] in 2005, looking at the *Wolbachia* endosymbionts of *Drosophila melanogaster* (*w*Mel) and *Brugia malayi* respectively (*w*Bm). Between 2005 to 2019, over 40

different *Wolbachia* genomes have been uploaded to the NCBI's Assembly database[137], spanning a range from insects through to nematodes. These 'new' *Wolbachia* genomes will have based much of their analysis on the genomes of *w*Mel and *w*Bm, yet these two genomes have received little in the way of reanalysis in light of modern knowledge, outside of automated algorithms such as NCBI's prokaryotic annotation pipeline[120]. The closest that is available is a recent *w*Bm resequencing effort using modern PacBio and Illumina-based techniques in 2019 by Lefoulon *et al.*[138]. Even then this study did not undertake a comprehensive reanalysis of *w*Bm's annotations, as it was not technically within the scope of investigation[120].

By contrast, the first genome sequence of a parasitic filarial nematode was that of *Brugia malayi*, sequenced in 2007 by Ghedin *et al.*[112]. A total of 12 filarial nematode genome sequences from the Onchocercidae family have been published between 2007 to 2019, several of which are known to harbour strains of *Wolbachia*.

In general, *Wolbachia* genomes across all supergroups range in size from 0.9 to 1.6 (Mbp) The mutualist *Wolbachia* of nematodes often exhibit smaller genome sizes within this range as compared to the parasitic *Wolbachia* of arthropods. As an example, *w*Bm was determined to have a genome size of 1.08 Mbp encoding 806 protein-coding genes on first publishing, whilst *w*Mel had a genome size of 1.26 Mb encoding 1,270 protein-coding genes. *Wolbachia* genomes in general are also comparatively smaller than what can be expected of bacteria in general, with the genome of the model bacteria *Escherichia coli* strain K-12 for example having a genome size of 4.64 Mbp[139]. This reduced genome size is a common feature amongst intracellular symbionts, such as other members of the Rickettsiales, as the biochemical pathways of the bacteria become degenerated and supplanted by the host over time[140,141]. Despite these reduced sizes, *Wolbachia* genomes in general have been noted to contain a large number of repeated sequences and mobile insertion sequence elements, which can be grouped into separate 'families' by sequence similarity. *Wolbachia* of arthropods were noted as often having a higher abundance of these elements when compared to *Wolbachia* of nematodes[111,118,142]. The families of insertion sequences in particular also appear to be *Wolbachia*-specific, with the genome sequencing project of *Wolbachia* in *Culex pipiens* (*w*Pip) noting that there was little family overlap with *w*Mel[143].

In addition, the presence of an integrated bacteriophage of *Wolbachia*, known as Prophage WO has been observed within multiple *Wolbachia* of arthropods, but appears either degenerated, or absent, in *Wolbachia* of nematodes[111,142,144–146]. This corroborates previous

observations that have identified the presence of a bacteriophage in 1978[147]. Subsequent studies by Beckmann *et al.*[148] and LePage *et al.*[149] in 2017 within prophage WO identified the presence of an operon of two genes (*cidA*-*cidB*). The two studies convergently identified the genes as acting together to form a bacterial toxin-antidote system to induce cytoplasmic incompatibility phenotypes in arthropods. Interestingly, this operon within phage WO has been noted to be degenerated in non-cytoplasmic incompatibility inducing strains of *Wolbachia*[150], such as *Wolbachia* of nematodes[111].

Genome sequencing of *Wolbachia* from different hosts have either not mentioned, or have been unable to find, evidence of flagellar, fimbrial, or pili genes that may be responsible for *Wolbachia* motility within the host[111]. It has instead been predicted that they migrate via co-opting of host actin filaments[111,151,152], which is a feature that appears conserved amongst the Rickettsiales[153,154]. In addition, while *Wolbachia* genomes generally maintain the genes required for the synthesis of lipid II- the major precursor monomer of peptidoglycan for cell walls, the pathways for synthesis of additional cell wall components D-alanine and D-glutamate appear absent. Additional enzymes that are used to catalyze the cross-linking of the cell wall are also absent in multiple *Wolbachia* genomes, particularly that of nematodes[111]. These truncated pathways go some way to addressing previous observations that the cell wall structure of *Wolbachia* is atypical[67].

Unusual for a member of the Rickettsiae, the genomes of all sequenced *Wolbachia* to date appear to conserve the pentose phosphate pathway, followed by the full *de-novo* biosynthetic pathway for purine and pyrimidine nucleotides[111,118,155]. Certain strains of *Wolbachia* also maintain the pathway for further refinement into the enzyme cofactors riboflavin and flavin adenine dinuleotide (FAD). Following on from this, the pentose phosphate pathway utilises intermediates from the glycolysis pathway of *Wolbachia*, which has been noted to be incomplete in many sequenced *Wolbachia* to date, both in arthropods and nematodes[111,118,136,142]. Specifically, *Wolbachia* lack the genes for direct uptake of extracellular glucose for entry into the glycolysis pathway (involving phosphoglucomutase and glucose-6-phosphate isomerase), and thus must rely on an alternative carbon source for feeding metabolic pathways, or ATP requirements. Hypotheses have focused on the potential for gluconeogenesis to remedy this gap, primarily due to the absence of enzymes catalysing irreversible glycolysis reactions- 6-phosphofructokinase, and pyruvate kinase[23,156]. In the place of pyruvate kinase, is a pyruvate phosphate dikinase (PPDK) enzyme which, while able to act in both the glycolytic and gluconeogenic directions, has been predicted to operate primarily in the gluconeogenic direction within *Wolbachia*[23,156]. The existence of PPDK in

Wolbachia has received some interest as a possible target for exploitation, as it is only known to be present in certain plants, bacteria, and archaea[157,158].

Finally, all *Wolbachia* genomes have been noted to encode components for the Type IV secretion system, a common transmembrane structure that is used by bacteria for transfer of material to a target. Commonly associated with virulence factors or transfer of DNA sequences as part of recombination, it could also be utilised by *Wolbachia* for transport of molecules to its nematode host. Type IV secretion system structures have also been directly observed on the surfaces of *w*Bm by immuno-transmission electron microscopy[159].

The first parasitic filarial nematode genome to be sequenced was *B. malayi*, first reported in 2004[160], then fully analysed in 2007, by Ghedin *et al.*[112], revealing a length slightly under 94 Mbp and containing 11,472 genes (NCBI Bioproject PRJNA27801). This is in comparison to the free-living nematode *Caenorhabditis elegans* which, according to the latest genome sequencing project in 2019, has a genome size of just over 100 Mbp, and encodes 28,416 genes (NCBI Bioproject PRJNA13758, Assembly GCA_000002985.3).

Since then, 12 different parasitic nematodes of the Onchocercidae family have been sequenced (**Table 1-1**), 8 of which have been published recently by the International Helminth Genomes Consortium[161]. The genomes of these 12 parasitic nematodes range in size from 77 Mbp (*Wuchereria bancrofti*) to 96.4 Mbp (*O. volvulus*), encoding between 10,397 (*Acanthocheilonema viteae*) to 16,117 genes (*Onchocerca flexuosa*).

The presence of *Wolbachia* DNA sequences within several filarial nematode genomes has also been identified in several studies[162], starting in 2007 by Hotopp *et al.* who observed a large number of *Wolbachia*-like DNA fragments within the *B. malayi* genome that was indicative of lateral gene transfer events[163]. But perhaps more interestingly, McNulty *et al.*[164] in 2010 identified the presence of *Wolbachia* sequences within the filarial parasites *Acanthocheilonema vitae* (a parasite of rodents) and *Onchocerca flexuosa* (a parasite of deer), with neither of these filarial nematodes known to harbour a *Wolbachia* endosymbiont[30,59,165]. After BLAST searches, McNulty *et al.* identified 49 and 114 *Wolbachia*-like DNA sequences within the two nematodes respectively, with 40 and 104 of the genes having identifiable homologues in *w*Bm[164]. As the study did not undergo comprehensive sequencing and analysis of the genomes of the two nematodes, the number of *Wolbachia*-like DNA sequences that have been integrated into these nematodes could be higher[164]. Regardless, the existence of these DNA fragments appears to be most congruent to the hypothesis that the *Wolbachia* endosymbiont has been lost over the course of evolution in

some species of filarial nematodes[23,164]. Whether this contributed to the eventual removal of *Wolbachia* in these nematode species remains to be elucidated.

Nematode	Assembly release	NCBI Assembly ID	Genome size (Mbp)	Protein-coding genes	*Wolbachia* present?
Acanthocheilonema viteae	2012	GCA_900537255.1	77.4	10,138	No
Brugia malayi	2007	GCF_000002995.3	93.7	11,472	Yes
Brugia pahangi	2018	GCA_900618355.1	90.5	14,599	Yes
Brugia timori	2018	GCA_900618025.1	64.9	15,861	Yes
Dirofilaria immitis	2012	GCA_001077395.1	84.9	10,179	Yes
Elaeophora elaphi	2018	***N/A***	82.6	10,410	No
Litomosoides sigmodontis	2018	GCA_900537275.1	64.8	10,246	Yes
Loa loa	2013	GCA_000183805.3	91.4	15,440	No
Onchocerca flexuosa	2018	GCA_900618345.1	86.2	16,066	No
Onchocerca ochengi	2018	GCA_900537205.1	95.5	13,698	Yes
Onchocerca volvulus	2016	GCA_000499405.2	96.3	12,534	Yes
Wuchereria bancrofti	2018	GCA_900622535.1	77.0	13,058	Yes

Table 1-1: Sequenced filarial nematode genomes available on the NCBI Assembly database for genomes[137], data taken on 29th July 2019. *Note that the complete genome sequence of Elaephora elaphi has not been released to the NCBI Assembly database as of September 2019, but has an associated publication available[161]*

1.3.5 Comparative genomics of *Wolbachia* and filarial nematodes

The essentiality of the *Wolbachia* endosymbiont for their nematode host's survival has long been a source of interest to the research community as a viable drug target for treatment options, as discussed earlier. Similarly, this same interest has extended into attempts to investigate the basic biology of symbiosis that exists between the two organisms. Initial comparisons between the genomes of *w*Bm and its nematode host revealed the presence of pathways that were intact in one, but missing in the other. This includes pathways for *de-novo* biosynthesis of important organic molecules or co-factors, such as purines, pyrimidines and haem[111], which are found within the *w*Bm endosymbiont but not within the nematode host. Similarly, *w*Bm was noted to maintain intact pathways for the *de-novo* biosynthesis of flavine adenine dinucleotide (FAD), another important co-factor. In turn, the nematode host maintains a complete glycolysis pathway for the import of extracellular glucose and conversion to intermediates for the Tricarboxylic acid (TCA) cycle, whilst the endosymbiont lacks the ability to import such extracellular glucose[111,118,136,142]. Instead, it has been hypothesised that *w*Bm utilises host pyruvate for the TCA cycle and energy production, or its unique PPDK enzyme in a gluconeogenic fashion for the synthesis of metabolites[23,111,156]. In

addition, the nematode host maintains genes for the biosynthesis of folate, as well as multiple amino acids, which the *Wolbachia* endosymbiont lacks[111].

Interestingly, analysis of the *Wolbachia* endosymbiont of *O. ochengi* (*w*Oo) in 2012[142], highlighted a lack of the riboflavin biosynthetic pathway in its entirety- a trait shared with the closely related *Wolbachia* endosymbiont of *O. volvulus*, sequenced in 2016[166], and a contrast to *w*Bm, which maintains the pathway[111]. Instead, these pathways appear to be at least partially present within the *Onchocerca* filarial hosts, a further contrast from the system in *B. malayi*. What is consistent between the two systems however, is the conservation of the *de-novo* nucleotide biosynthetic, as well as haem biosynthetic, pathway within the *Wolbachia* endosymbionts, and absence of the corresponding pathways within the nematode host[142,166]. In addition, both *w*Oo and *w*Ov appear to maintain genes for the *de-novo* biosynthesis of folate, which may be partially present in *O. ochengi*, but not in *O. volvulus*, marking a further contrast to the *w*Bm and *B. malayi* system.

Further analysis of *Dirofilaria immitis* and its *Wolbachia* endosymbiont (*w*Di) was performed in 2012 by Godel *et al.*[155]. Like other *Wolbachia* endosymbionts of nematodes as mentioned, here the authors observe a conservation of the *de-novo* biosynthetic pathways of haem and nucleotides within *w*Di. Much like *w*Oo, and unlike *w*Bm, the authors also noted the presence of genes for the *de-novo* folate biosynthetic pathway within *w*Di.

It is interesting to note that despite the individual differences that exist between the nematode-*Wolbachia* systems, all nematode *Wolbachia* appear to conserve the haem and nucleotide biosynthetic pathways, which their nematode hosts universally lack. Indeed, *B. malayi* has previously been demonstrated to be able to uptake exogenous haem using the transporter *Bm*HRG-1, demonstrated via utilisation of modified yeast cells that require exogenous haem to survive[167]. The same study also showed that RNA-interference studies targeting *Bm*HRG-1 resulted in a block in adult female nematode embryogenesis within 2 days- a strikingly similar phenotype to that seen in *Wolbachia*-depleted nematodes[167]. It is likely that the nematode acquires this haem from the *Wolbachia* endosymbiont, due to the nematode's lack of a developed digestive tract in most life cycle stages, as well as the difficulty of acquiring exogenous haem from the environments the nematode inhabits during its life cycle[167].

1.3.6 Application of Transcriptomics to study *Wolbachia* interactions with their filarial hosts

As of 2019, there have been 5 studies that have looked at the transcriptome of *Wolbachia* concurrently with their nematode hosts, namely *B. malayi*[168,169] (investigating the transcriptomes of L1 to adult male/female nematodes), *D. immitis*[170,171] (investigating the transcriptome of different life cycle stages or different adult nematode tissues) and *O. ochengi*[142] (investigating the transcriptome of different adult nematode tissues). These studies have largely focused on investigating later life cycle stages of the nematode (L3 to adults), or specific tissues of adults, with an exception of the study performed by Chung *et al.* in 2018[168]. The authors of this study have published, but has yet to analyse, the transcriptome from L1 through to reproductive adults.

Generally, the analysis of these transcriptome studies has corroborated hypotheses that were generated from analysis of the corresponding genome sequencing projects, and the importance of certain pathways (such as *de-novo* haem or purine/pyrimidine biosynthesis) in the symbiotic relationship. In addition, it has been observed that the transcriptomes of both the *Wolbachia* endosymbiont and their nematode hosts appear to be highly coordinated depending on the life cycle stage of the nematode and to a more limited extent the various tissues of the nematode[142,169–171]. For instance, within the *B. malayi* system it was observed that the DNA replication machinery, components of the haem biosynthetic pathway, as well as membrane transport functions, such as the Type IV secretion system, show a high basal level of transcriptional activity in *w*Bm[169]. During later, reproductively active nematode life cycle stages, *w*Bm glycolysis/gluconeogenesis pathways were upregulated when compared to developing life cycle stages- this was accompanied by *B. malayi* genes responsible for glycogen catabolism also being upregulated, indicating *w*Bm's dependence on the host for energy production[169]. In developing female nematodes, certain components of the nucleotide biosynthesis pathways in *w*Bm were also upregulated, as were additional genes for energy production and chaperone functions, with these chaperones known to be expressed in response to oxidative stress or heat shock[169].

Within *O. ochengi*, it was noted that *w*Oo within the nematode germline exhibited upregulation of components of the DNA replication, as well as translational machinery, likely indicating that *w*Oo is rapidly replicating so as to invade host oogonia and facilitate transfer to the next generation[142]. By contrast, *w*Oo within the nematode soma exhibited upregulated membrane transport machinery, such as metal ion transporters, components of the oxidative

phosphorylation pathway, Type IV secretion system components, as well as the Sec secretion components[142].

Within the *D. immitis* system, the haem biosynthetic pathway of *w*Di was noted to be highly upregulated during the mf life cycle stages, with other studied life cycle stages showing a low level of transcriptional activity in this pathway[170]. The authors postulate that haem biosynthesis might be highly upregulated during mf stages as a form of 'stockpiling'[170], and preparation for *D. immitis*' development within the mosquito vector, similar to *Plasmodium* parasite species[172]. This profile is replicated for the purine and pyrimidine biosynthetic pathways, as well as membrane transport mechanisms, although *w*Di showed upregulation in both mf as well as the L4 life cycle stages[170]. These observations suggest stage-specific expression and supplementation by *w*Di for its nematode host[170].

1.4 Project aims

Despite the existing knowledge gained from analysing and comparing the genomes and transcriptomes of *Wolbachia* and their nematode hosts, the fact remains that no studies have yet to fully investigate the transcriptome of developing nematode larvae in detail. More specifically, there is an as-yet unelucidated role that *Wolbachia* appears to play in allowing transmission of the nematode to their vector hosts. In addition, *Wolbachia*'s role during early L3 to L4 nematode development shortly after infection of the final host, as well as the processes that allows its rapid population expansion, has yet to be elucidated. And underpinning this all, is the genome sequence and annotation of *w*Bm that, despite being the basis for over 40 different *Wolbachia* genome sequencing projects, is nearing 20 years of age with only automated curation by NCBI[120], and lacking any human curation using recent technological advances.

As such, the research questions this thesis aims to address can be summarised as follows:

1. With the benefit of new techniques, knowledge, and human curation, does the *w*Bm genome contain any previously unannotated genes that can reveal new biology and/or interactions with its nematode host?
2. Utilising the L3-L4 developmental transcriptome of *B. malayi* and its *w*Bm host, what are the interactions between host and endosymbiont during this key developmental stage, and are there any metabolic pathways that can explain *w*Bm's population growth during these stages?
3. Utilising antibiotic-treated and untreated *B. malayi* microfilariae, what effects does *w*Bm depletion have on *B. malayi* microfilariae's ability to infect the vector host and,

utilising transcriptomic data, are there any distinct pathways, complexes, or enzymes that could explain any observations made?

These research questions aim to provide a better understanding of the relationship *w*Bm may have with its nematode host. At the genomic level, this could reveal new biology that may be applicable to other *Wolbachia* genome sequences that have been released since 2004-2005. At the phenotypic level, identification of genes, pathways, or complexes that play a role in symbiosis could be relevant to other *Wolbachia*-nematode symbiotic relationships, but also may have implications for wider disease, or potentially vector, control strategies.

Chapter 2 Reannotation of the *w*Bm genome

2.1 Introduction

Development of technologies within the last decade in the form of massively parallel sequencing by companies such as Illumina or PacBio[116] has meant that an increasing number of genomes are being sequenced and entering the public domain. However, although sequencing technologies have advanced, annotation and analysis of these resulting genomes has lagged in comparison, with automated strategies for genome annotation becoming the norm due to the sheer amount of data generated. Automated strategies are typically based upon pre-set algorithms trained on model organisms, and require frequent updates to keep up to date with the generation of new biological knowledge.

The developments of such high-throughput genome sequencing technology have also lead to the development and proliferation of RNA-sequencing technology[173]. This is a powerful technique that can be applied to study the transcriptome of an organism, generating knowledge as to what genes are expressed at certain time-points, developmental stages, or environmental pressures. In addition, the technique's ability to accurately identify gene expression in organisms lends itself well for annotation projects[173]. Specifically, if computationally predicted genes can be shown to have active transcription and gene expression by RNA-sequencing, the validity of the predicted gene coordinates would be strengthened significantly. As a result of this, RNA-sequencing techniques have been applied in recent years to validate newly sequenced genomes, as well as reannotate genomes that were sequenced before the technology became widely available. Two recent examples of such reannotation work were performed by Tran *et al.* on the fungal dermatophyte *Arthroderma benhamiae*[174], and Yang *et al.* on 8 different *Drosophila* species[175]. RNA-sequencing helped Tran *et al.* to identify modifications to 64% of the existing *A. benhamiae* genome, including annotation of 383 new genes and deletion of 708 genes[174], whilst Yang *et al.* identified between 1,200 to 3,800 new genes[175], as well as extension of the coding sequence of many existing genes, within the studied 8 *Drosophila* genomes.

In the case of the *Wolbachia* endosymbiont of both insects and parasitic filarial nematodes, *Wolbachia* of *Drosophila melanogaster* (*w*Mel) and *Brugia malayi* (*w*Bm) were the first two *Wolbachia* genomes to be sequenced in 2004[136] and 2005[111] respectively. This yielded two comparatively small genomes of 1,267,782 and 1,080,084 base-pairs respectively, with subsequent annotation of these genomes relying initially on several automated programs, including the ERGO software suite and GeneMarkS or GLIMMER programs[111,176,177], followed

by manual human curation. This predicted 1,195 and 805 protein-coding genes within the genomes of *w*Mel and *w*Bm respectively. By comparison, these genomes are significantly smaller than the genome of the model organism *Escherichia coli* K-12, first sequenced in 1997[139], which was determined to have genome length of 4,639,221 base-pairs and maintained over 4,000 genes.

Whilst the automated annotation programs used on the *Wolbachia* endosymbionts have received multiple updates to keep up with knowledge[178,179], they have not been re-run on these genomes to take advantage of these updates. For instance, the genome of *w*Bm has very recently been resequenced, but not reannotated, in 2019 using PacBio's Single Molecule Real-Time (SMRT) technology[138]. This resequencing effort has revealed a slightly larger genome than previously thought (1,080,939 vs 1,080,084) and identified a minimum of 18 single-nucleotide variants (SNVs) as compared to the original 2005 genome[138], with these 18 subsequently being experimentally validated. The authors do state however, that it is difficult to know if these SNVs are due to sequencing errors, or genuine biology arising from the 14 year-gap between sequencing efforts.

A second *Wolbachia* genome that has been resequenced in recent years is that of *Wolbachia* in *Aedes albopictus* (*wAlbB*)[118]. *A. albopictus* mosquitoes are known to be infected with two strains of *Wolbachia* from supergroups A and B. The resulting cell lines of these mosquitoes, such as Aa23, retain only supergroup B[118,180], and have been a useful model organism for studying host-endosymbiont interaction. First sequenced in 2012 by Mavingui *et al.*[181] using a 454 Titanium machine, this identified a genome length of 1,239,814 base-pairs and 1,209 protein coding genes, which is approximately in line for a *Wolbachia* endosymbiont of insects. In 2019, the genome of *w*AlbB was resequenced by Sinha *et al.*[118] utilising a combination of PacBio and Illumina sequencing technology. This identified a genome length of 1,484,007 base-pairs, with all of the 2012[181] genome mapping to this new assembly[118], making the genome of *w*AlbB one of the largest of insect *Wolbachia*. In addition, 1,205 protein coding genes were identified via the National Center for Biotechnology Information (NCBI) prokaryotic annotation pipeline[120]. This is an automated genome annotation pipeline that combines information from various sources, such as the wealth of annotation knowledge from NCBI's genome and short-read archive databases, as well as gene or pseudogene prediction tools such as GeneMarkS+[120].

These two genomes (of *w*Bm and *w*AlbB) are the only two *Wolbachia* genomes that have received any form of resequencing or reanalysis, despite over a decade of research into

Wolbachia endosymbionts. This lack of reannotation efforts to take advantage of new knowledge and updated algorithms poses a significant problem in understanding *Wolbachia* biology. Since the publication of genomes for *w*Bm in 2005, over 40 different *Wolbachia* genomes have been sequenced to various stages of completion (NCBI's Assembly database[137]), all of which will have been built upon the knowledge first imparted by the genomes of *w*Mel and *w*Bm. These 40 genomes represent only a small fraction of the diversity present in the *Wolbachia* genus, which can be subdivided into 16 supergroups, A through to Q[36], of which *w*Bm is a member of Supergroup D. This supergroup, as well as Supergroup C *Wolbachia*, contain the majority of *Wolbachia* that live in symbiosis with a nematode host, almost all of which show a mutualistic relationship.

The objective of this chapter is to undertake a comprehensive review and reannotation of the *w*Bm genome, as well as analysis of this genome in the context of other newly-sequenced *Wolbachia* genomes. This will be achieved utilising modern knowledge and technology, as well as transcriptomic data derived from *B. malayi* nematodes collected 14 days post-infection of Mongolian jirds, with additional RNA-sequence data from days 3, 7 and 11 serving as validation. The 'total' transcriptome from these time-points was first taken from the *B. malayi* host and partitioned into 2 datasets by alignment to their respective genomes[111,112]; **(i)** *B. malayi*-unique reads plus lateral gene transfer events, and **(ii)** *wBm*-unique reads. This is then followed by manual curation of the genome to identify possible coding regions, pseudogenes, or RNA-coding genes, before the addition of bioinformatic tools such as Blast2GO[182] to complete the reannotation of *w*Bm's genome (see **Figure 2-1** for workflow overview). This time-point was chosen for investigation due to its significance in the nematode and *w*Bm life cycle, occurring just after the L3-L4 moult as well as the rapid population growth of *w*Bm[15]. This period of population expansion would theoretically result in the expression of much of *w*Bm's genes, thus aiding reannotation efforts.

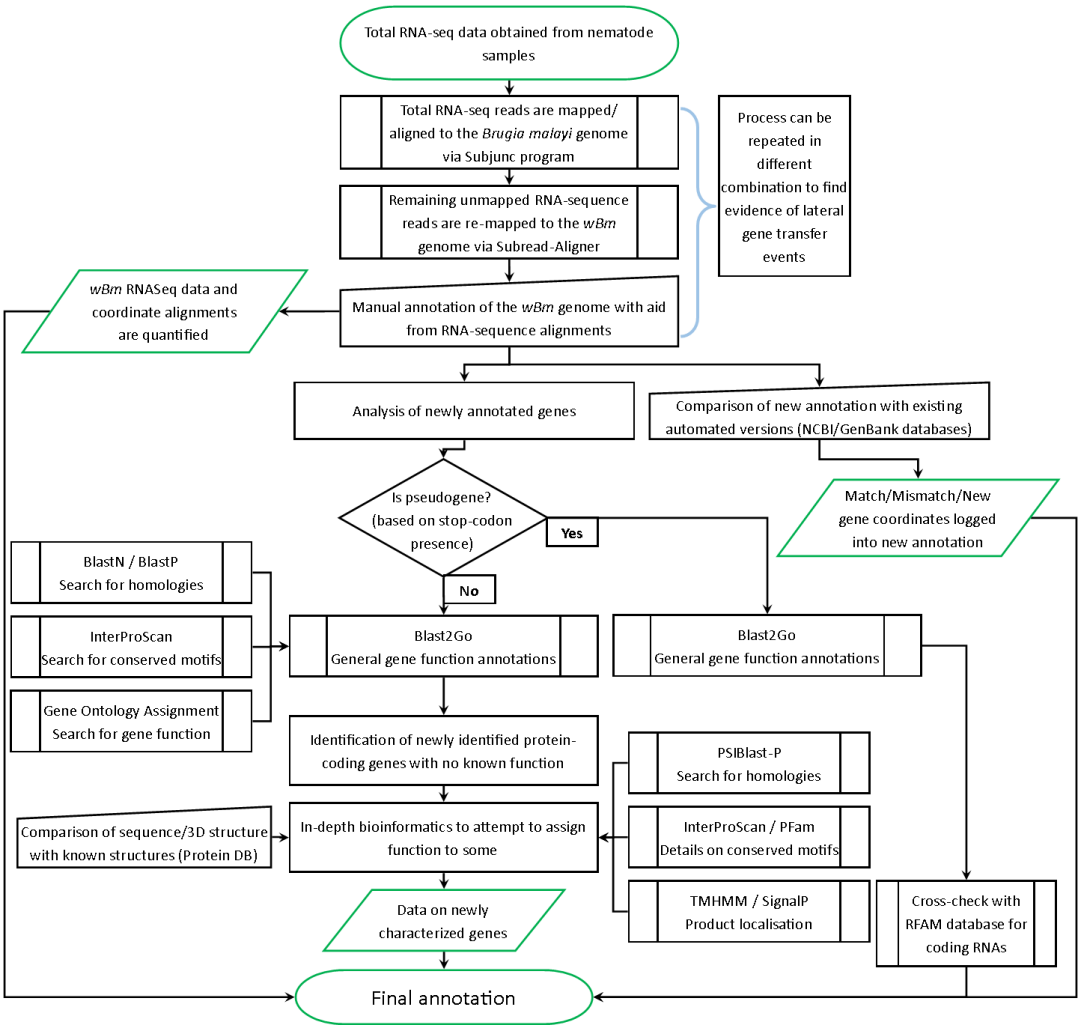


***Figure 2-1: Overview of experimental design for identifying and annotating new genes in* wBm.** *This comprises of RNA-sequencing and genome alignments, followed by manual curation and annotation of regions in the genome that show high expression, before characterization of newly-annotated regions for either protein-coding genes, pseudogenes, or RNA-coding genes. This characterization process utilizes a variety of tools*[122,131,182–187].

2.2 Results

2.2.1 RNA-Sequencing and Alignment of RNA-sequence data to the *w*Bm Genome

All RNA extraction work in this section was done by Dr Christina Bronowski, with processing and sequencing done by the Liverpool Centre for Genomic Resources

To identify unannotated regions of the *w*Bm genome, parasites were obtained from the Filariasis Research Reagent Resource (FR3) laboratory that spanned the first two weeks of nematode development within the Mongolian jird model (*Meriones unguiculatus*). This involved injection of infective L3-stage nematodes into the peritoneal cavity of the jird, and incubation for 3, 7, 11, and 14 days, before the nematodes were then recovered via necropsy. For both day 3 and day 7 time-points this yielded 2,000 L3 stage nematodes each, and days 11 and 14 yielded 1,000 L4 stage nematodes each. RNA was then collected via TriZol

extraction, depleted of rRNAs via Terminator Exonuclease, and sequenced using an Illumina GA-IIx machine. Terminator exonuclease treatment was used to remove rRNAs from both eukaryotic (*B. malayi*) and prokaryotic (*w*Bm) sources. RNA-sequencing from the day 14 time-point yielded a total of 80,258,564 trimmed unaligned reads across 2 separate, single-ended technical replicates and a final paired-end technical replicate. As mentioned previously, this study utilizes the day 14 time-point for reannotation efforts, whilst the remaining 3 time-points served as additional validation across an additional life-cycle, as well as to aid in the detection of single-nucleotide variants. The full results from all 4 time-points are described in **Chapter 3**.

To obtain RNA-sequencing data from the *w*Bm endosymbiont, we designed a bioinformatics pipeline that allowed us to extract the *w*Bm transcriptome from the significantly more abundant *Brugia malayi* transcriptome. This is first done by aligning the total transcriptome dataset to the *B. malayi* genome using a splice-aware aligner, before extracting unmapped reads and realigning these to the *w*Bm genome. This operation can be run in the reverse order, and the resulting alignments can be compared to observe any potential cases of lateral gene transfer, which are known to occur in such *Wolbachia*-host symbiotic relationships[162,163].

Using Subread-Aligner[188], 88.45% of the first reads were successfully mapped to the *Brugia malayi* genome. The remaining 9,273,693 unmapped reads were then remapped to *wBm*, resulting in 26.91% reads successfully mapped (3.11% of total reads sequenced mapped to *wBm*). The remaining sequence reads that failed to map to either genome were discarded. The mapped reads obtained from this alignment were used to determine length normalised expression levels for genes in *wBm*'s genome. These were calculated by Cufflinks[189] as 'Fragments Per Kilobase of exon per Million mapped reads' (FPKM) before aligned reads were visualised using Artemis[190,191]. Aligned RNA-sequence reads were used to confirm the position of existing genome annotations[111], and used to identify potential sites for new annotations to be added manually. In brief, regions of the genome were designated as Regions of Interest (ROI) if they contained significant RNA-sequence alignments relative to background, with such sites being identifiable by the presence of large 'stacks' of RNA-sequence reads aligned to the genome (**Figure 2-2**). Relative expression levels of all ROI's were compared by the generation of a density plot using FPKM values, with a 5% confidence interval (equalling an FPKM of 5.4) in the dataset being established, and newly identified ROI being considered significant if FPKMs exceeded this threshold (**Figure 2-3**).

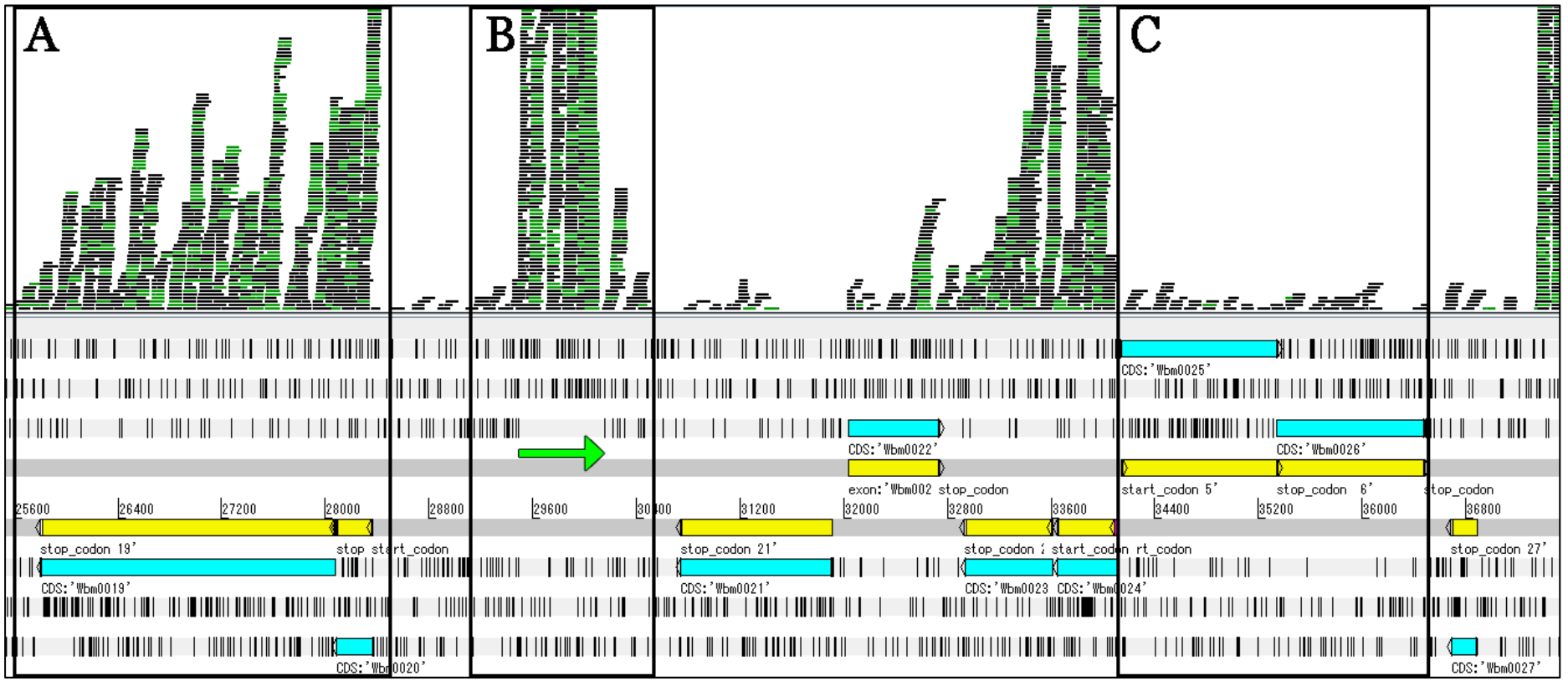


***Figure 2-2**: **Example output of the Artemis genome viewer, focusing on the* wBm *genome and RNA-sequence data generated from this study.** Note the top half of the image which shows a visualization of aligned RNA-sequence reads, illustrated as black and green 'stacks'. Also note the bottom half of the image which shows the 6 different reading frames in either the forward or reverse direction. Stop codons within these reading frames are identified by black bars, with annotated genes identified by Foster et al. 2005 illustrated as blue (Coding Sequences) and yellow (exon) bars located within regions that have no stop codons. In box **(A)**, note the presence of two annotated genes identified by Foster et al. and the large RNA stacks above both (*Wbm0019 *and* Wbm0020*), indicating high transcriptional expression. In box **(B)**, note the presence of a large RNA stack, but no annotated regions. Instead, in this box note the presence of a genome region with no stop codons (**green arrow**). In box **(C)**, note the presence of two genes that have comparatively small RNA stacks (*Wbm0025 *and* Wbm0026*), indicating low transcriptional expression.*

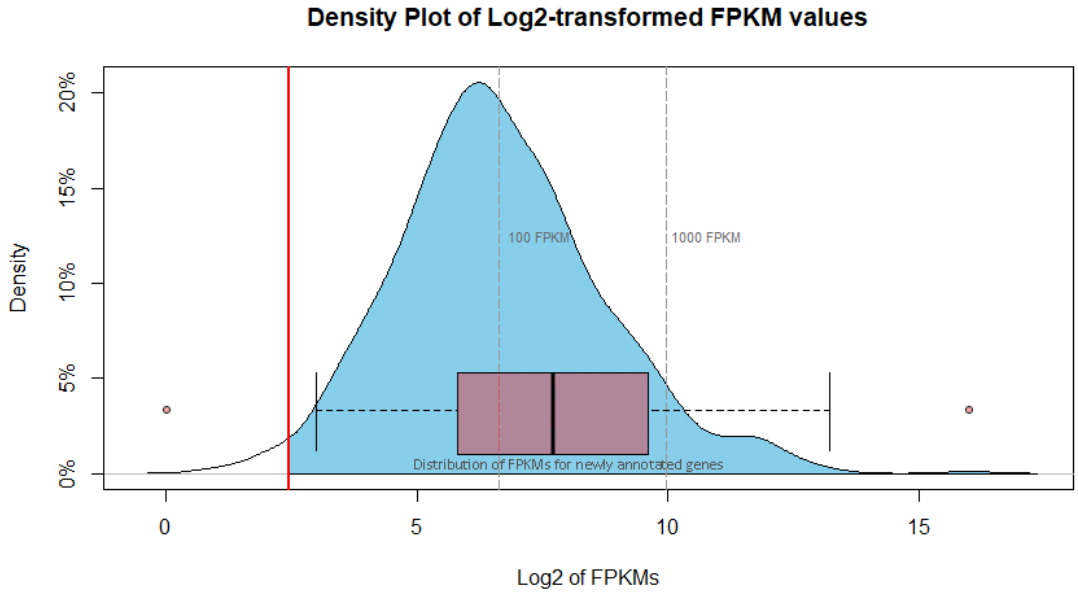


Figure 2-3: Density plot of FPKM data. *Note the labelled, solid red line which signifies the cut-off point set in this study, above which a gene is considered expressed, as represented by the area of the density plot shaded in blue. The red line signifies the 5% threshold used in this study (5.4 FPKM), whilst the grey dashed lines and labels show reference points for 100 and 1,000 FPKM on the log_2 transformed scale. The overlaid boxplot indicates the distribution of FPKMs for genes that were newly annotated during this study. Note that, with the exception of 1 gene, all of these newly-annotated genes exceeded the 5% FPKM threshold and have comparable FPKMs to existing genes.*

From this distribution of FPKMs, we determined the median and mean FPKMs for the protein-coding regions predicted by Foster *et al.* in 2005[111] at 81 and 278 FPKM respectively. In addition to this, there were several regions of the *w*Bm genome that lacked annotation but had observable RNA-sequence stacks that were comparable in size to existing protein-coding regions (**Figure 2-3**, boxplot indicates FPKM distribution of all newly identified genes). Assuming that the majority of original annotations in the *w*Bm genome determined by Foster *et al.*[111] are correct, this indicated the presence of several new genes, pseudogenes, and non-standard transcriptomic events. Both original and new gene predictions were combined as a single annotation file in General File Format (GFF, **A 1-1**).

2.2.2 Overview of newly identified putative coding genes

A total of 1,048 ROI, which excludes *wBm*'s previously annotated 34 tRNAs and 3 rRNAs were identified in the analysis. This comprises 959 ROI that met the FPKM threshold (**Figure 2-4A**, 'Expressed RNA' circle) and a further 89 newly identified pseudogenes, or genes from the original genome annotation that failed to meet the FPKM threshold (**Figure 2-4A**, red and green segments). ROI were designated as protein coding if they had expressed RNA and at least one of the following: **(i)** a clear open reading frame of at least 300 nucleotides (**Figure 2-4A**, intersect with 'Clear Open Reading Frame' circle, total 784) **(ii)** near-full length

homology to an entry with known function in NCBI protein databases (**Figure 2-4A**, intersect with 'Related Annotations Exist' circle, total 843) or **(iii)** both (**Figure 2-4A**, yellow segment, total 697). Regions were also designated as pseudogenes if the reading frame was interrupted, or if high FPKMs were noted in a region despite a lack of open reading frame (**Figure 2-4A**, union of green and blue segment, total 222). This final category does include 3 predicted functional-RNA genes, and 5 cases of potential stop-codon readthrough or ribosomal frame-shifting.

This read-mapping and manual inspection of gene coordinates in *w*Bm was repeated using additional RNA-sequence datasets obtained from existing databases[192,193] (NCBI bioproject number PRJNA294264, NCBI Gene Expression Omnibus number GSE93139). This data originates from different life cycle stages of the *B. malayi* host, where the *w*Bm endosymbiont will have a different transcriptome profile as compared to the L3 developmental stages.

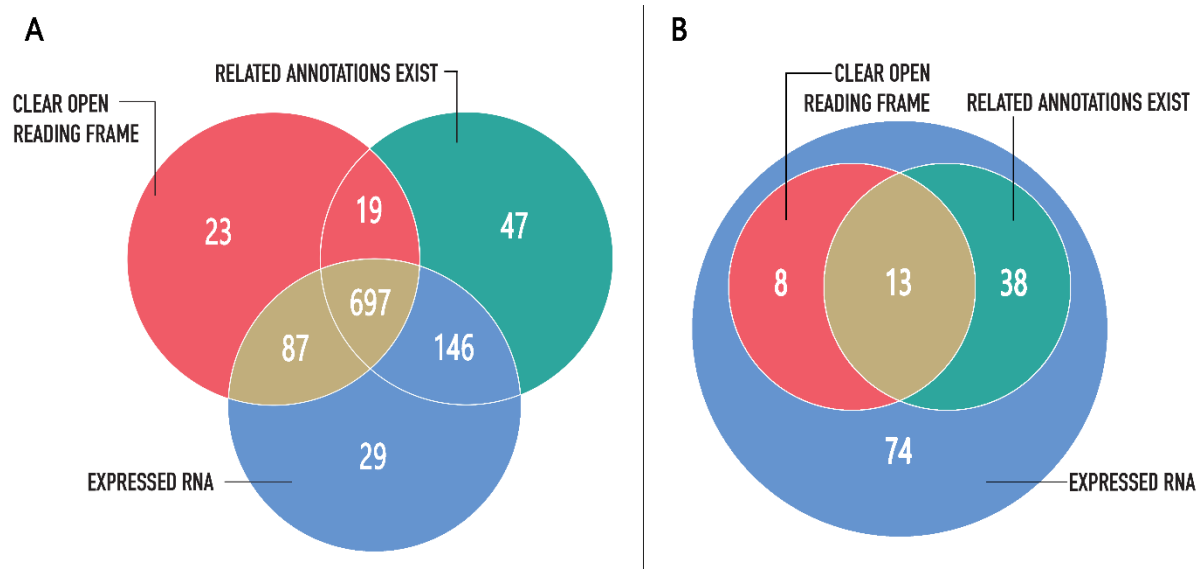


Figure 2-4: Venn diagrams to show distribution of identified genes in* wBm*, (excluding tRNAs and rRNAs). Each gene was categorised as having (i) a clear open reading frame (red) (ii) evidence of expression by RNASeq (blue) or (iii) an annotated homologue in the NCBI protein database (green). ***Section A*** *represents an overview of the complete* wBm *genome annotation, combining new genes predicted from RNASeq expression evidence with the 2005 annotations.* ***Section B*** *represents an overview of the newly predicted genes and pseudogenes identified using expression data from this study.*

This analysis identified a total of 133 new potential genes in *wBm*, (**Figure 2-4B**). Initial analysis identified 20 genes that were considered as potentially protein-coding due to having a clear ORF and meeting the 5% threshold to be considered as expressed. Of these, 12 were found to have sequence relationships to genes or transcripts in a database that was indicative of known function, or had general structural motifs that can be ascribed to function (**Table 2-1**, **Figure 2-4B**, yellow segment). Conversely, the remaining 8 had sequence relationships

to genes or transcripts which were described as 'hypothetical' or 'putative' in existing databases (**Figure 2-4B,** red segment).

Following manual review, we elected to include one additional gene (wBmNew0012) with evidence of an ORF of 396 nucleotides despite not meeting the 5% FPKM threshold, making a total of 21 newly discovered genes. This gene showed between 46 to 60% identity over approximately 40% of query coverage to an Ankyrin-repeat containing protein that had been identified in other sequenced *Wolbachia* genomes, such as *Wolbachia* of *Cimex lectularius* (bedbug) or other arthropods (**Table 2-1**, **Figure 2-4B**, red and yellow segments).

In addition to these coding genes, 3 additional functional-RNA genes were identified which did not have a clear ORF but had appropriate annotations when compared against the RFAM database[184] (**Figure 2-4B**, included as part of green segment).

Of the 15 genes with known function (including the 3 functional RNA genes), 3 encoded for separate copies of a VirB2 subunit, 1 encoded for VirB7 subunit, and one encoded a SecE subunit. These 5 transport-related genes are comparatively small in size (< 400 nucleotides, **Table 2-1**), which makes it more likely for automated detection methods to misidentify them.

The VirB2 subunits (*wBmNew0003, wBmNew0008, wBmNew0022*), while small, act as an integral component of the Type IV secretion system, and at least two of the three newly identified copies are predicted to have full length open reading frames and significant transcript expression levels (130 FPKM minimum). These subunits are responsible for constructing the connecting pilus that allows for substrate transfer between the originating bacterium and its target[194,195] (**Figure 2-5**).

The newly identified VirB7 subunit (*wBmNew0002*) is another integral component of the Type IV secretion system. This subunit shows high transcript expression levels when compared to general FPKM distribution (385 FPKM), and functions as part of the Type IV secretion system's core complex[194,195]. Together with VirB9 and VirB10 (which were already identified in the 2005 annotation), this complex forms the main pore channel that spans the transmembrane region, as well as both the inner and outer membranes[194,195]. VirB7's specific role appears to keep this core complex stable by fusing VirB9 and VirB10 together, whilst localised at the outer membrane portion of the complex[194,195] (**Figure 2-5**).

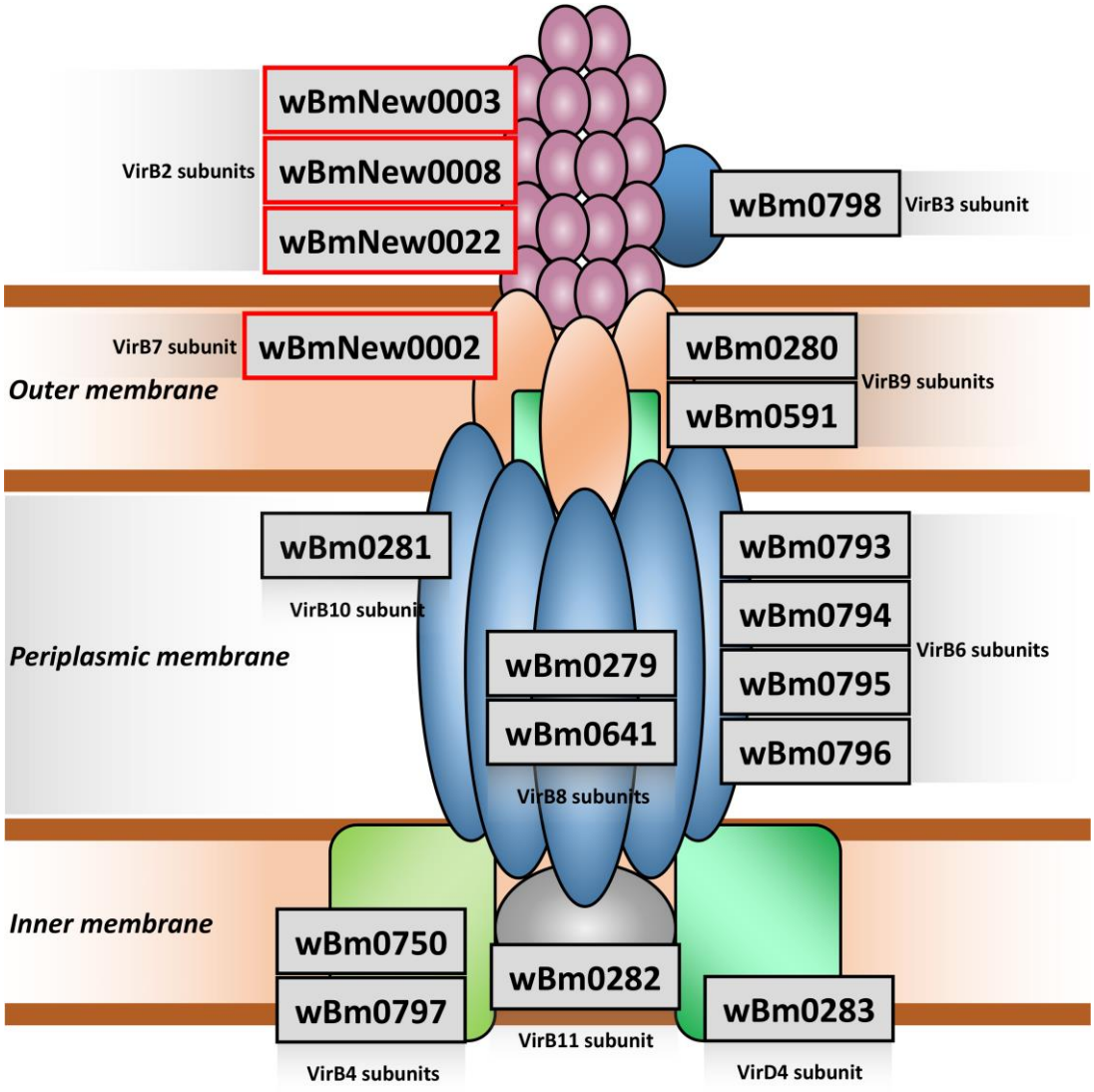


Figure 2-5: Representation of the Type IV secretion system in wBm, with newly identified genes from this study highlighted in red.

The SecE subunit (*wBmNew0020*) again shows high transcript expression levels when compared to the general distribution (215 FPKM, **Figure 2-3**), and is a critical stabilising member of the heterotrimeric Sec-translocase system complex[196,197], which is a major pathway for the translocation of proteins across bacterial membranes[198]. Translocated proteins can function within *wBm*'s periplasm, be embedded into the membrane, or excreted into the extracellular space and thus into the nematode host. Together, these 5 newly identified genes may have implications for the nutrient exchange machinery between *wBm* and its nematode host.

Outside of these transport-related genes, the gene *wBmNew0009* was noted to have a 229 amino acid sequence with 100% identity to the gene *wBm0395*, which encodes for the E1 component of the 2-oxoglutarate dehydrogenase complex and has a full length of 887 amino

acids. This 100% identity implies *wBmNew0009* may be the result of a truncated, gene duplication event. The newly identified gene was noted to have an FPKM value of 166.6, which is comparable to the full-length gene's FPKM of 255.1 indicating that both genes show expression. Investigation of *wBmNew0009* did, however, identify several reads that overlapped both the open reading frame of the gene as well as the surrounding intronic regions. This indicates that this duplicated gene may be expressing transcripts directly and that reads from *wBm0395* are not double mapping.

These new coding features have been combined with the existing 805 protein-coding and 37 RNA-coding regions identified by Foster *et al.* in 2005[111] into one GFF annotation file (Gene Feature Format, version 2.2), thus describing a combined total of 866 genes and RNA coding regions. All regions newly identified from this study have been designated with the prefix '*wBmNew*' in the GFF annotation files (**A 1-1**). A second GFF file contains information on the 110 pseudogenes identified in this study, the 104 pseudogenes identified by Foster *et al.*, as well as 5 instances of non-standard translational events, for a total of 219 pseudogenes (**A 1-2**). Additionally, 25 genes identified by the Foster *et al.* annotation were noted to have RNA-sequence alignments outside of their annotation coordinates, with BLAST searches of the flanking regions identifying full, or near-full length hits, providing evidence for the modification of the 25 gene's start codons (**Table 2-2**). These regions that had modified reading frames have had a note included in the appropriate GFF file (**A 1-1**).

Sequence ID	Length (nuc)	Database Description
wBmNew0001	648	**Database Match – Unknown function, membrane-bound**; 74% similarity to gene in *Wolbachia* of *Onchocerca ochengi*
wBmNew0002	168	**Type IV secretion system protein VirB7**; 79% similarity to gene in *Wolbachia* of *Onchocerca ochengi*
wBmNew0003	276	**Type IV secretion system, TrbC/VirB2 family protein**; 72% similarity to gene in *Wolbachia* of various *Drosophila*
wBmNew0004	267	**Database Match – Unknown function, membrane-bound**; 72% similarity to gene in *Wolbachia* of various *Drosophila*
wBmNew0005	627	**Oxidoreductase NAD-binding domain protein**; 94% similarity to gene in *Wolbachia* of *Cimex lectularius*
wBmNew0006	510	**Predicted membrane protein**; 82% similarity to gene in *Wolbachia* of *Cimex lectularius*
wBmNew0007	312	**Predicted Lipoprotein**; 57% similarity to gene in *Wolbachia* of *Culex quinquefasciatus*
wBmNew0008	339	**Type IV secretion system, TrbC/VirB2 family protein**; 87% similarity to gene in *Wolbachia* of various *Drosophila*
wBmNew0009	687	**2-oxoglutarate dehydrogenase subunit E1**; 94% similarity to gene in *Wolbachia* of various *Drosophila*
wBmNew0010	153	**6S / SsrS RNA**; RFAM e-value of *6.7e-12*
wBmNew0011	396	**Transposase DDE Domain**; 65% similarity to genes in *Flavobacterium*, *Capnocytophagia*, and other gram-negative bacteria
wBmNew0012	396	**Ankyrin repeat containing protein**; 75% similarity to gene in *Wolbachia* of *Cimex lectularius*
wBmNew0013	198	**Database Match – Unknown Function, membrane-bound**; 80.7% similarity to gene in *Wolbachia* of various *Drosophila*
wBmNew0014	252	**Database Match – Unknown Function**; 89% similarity to gene in *Wolbachia* of *Wuchereria bancrofti*
wBmNew0015	349	**Bacterial RNAse P Class A**; RFAM e-value of *8e-94*
wBmNew0016	480	**Phosphatidylglycerophosphatase A**; 94% similarity to gene in *Wolbachia* of *Cimex lectularius*
wBmNew0017	246	**Database Match – Unknown Function, membrane-bound**; 93% similarity to gene in *Wolbachia* of *Wuchereria bancrofti*
wBmNew0018	471	**Database Match – Unknown Function, no identifiable domains**; 71% similarity to gene in *Wolbachia* of *Dactylopus coccus*
wBmNew0019	285	**Alpha transfer-messenger RNA**; RFAM e-value of *3e-59*
wBmNew0020	198	**Preprotein translocase subunit SecE**; 95.5% similarity to gene in *Wolbachia* of various *Drosophila*
wBmNew0021	372	**Ankyrin repeat domain containing protein**; 66% similarity to gene in
wBmNew0022	327	**Type IV secretion system, TrbC/VirB2 family protein**; 87% similarity to gene in *Wolbachia* of *Muscidufax uniraptor*
wBmNew0023	222	**Database Match – Unknown Function, no identifiable domains**; 100% similarity to gene in *Wolbachia* of *Wuchereria bancrofti*
wBmNew0024	156	**No Database Matches**

Table 2-1: List of newly identified genes from this study. *Note the presence of multiple Type IV secretion system components and functional RNAs that were identified from this study.*

Gene ID	Strand	Product	Modification
WBm0011	Antisense	Permease of the major facilitator superfamily	Start position elongated 45 nucleotides sense direction
WBm0013	Sense	Predicted protein	Start position elongated 102 nucleotides antisense direction
WBm0117	Antisense	Lipoprotein	Start position elongated 42 nucleotides sense direction
WBm0142	Sense	Predicted phosphatase	Start position elongated 99 nucleotides antisense direction
WBm0208	Sense	Predicted membrane protein	Start position elongated 48 nucleotides antisense direction
WBm0216	Antisense	ABC-type protease secretion system, ATPase and permease component, AprD	Start position elongated 120 nucleotides sense direction
WBm0257	Sense	Mismatch repair ATPase, MutS family	Start position elongated 45 nucleotides antisense direction
WBm0260	Sense	Dioxygenases related to 2-nitropropane dioxygenase	Start position elongated 21 nucleotides antisense direction
WBm0285	Antisense	Methylase of polypeptide chain release factor	Start position shortened 96 nucleotides antisense direction
WBm0421	Antisense	Multisubunit Na+/H+ antiporter, MnhE subunit	Start position elongated 150 nucleotides sense direction
WBm0424	Antisense	Na+/alanine symporter	Start position elongated 90 nucleotides sense direction
WBm0426	Antisense	Phosphatidylserine decarboxylase	Start position elongated 39 nucleotides sense direction
WBm0439	Antisense	Oligoketide cyclase/lipid transport protei	Start position shortened 111 nucleotides sense direction
WBm0481	Antisense	Membrane-bound protoheme IX biogenesis protein, HemY	Start position elongated 57 nucleotides sense direction
WBm0546	Sense	Predicted esterase	Start position elongated 33 nucleotides antisense direction
WBm0559	Antisense	Dihydrolipoamide dehydrogenase E3 component	Start position shortened 51 nucleotides antisense direction
WBm0584	Antisense	Ubiquinone biosynthesis protein COQ7	Start position elongated 51 nucleotides sense direction
WBm0598	Sense	Protein containing GNAT family acetyltransferase domain	Start position elongated 90 nucleotides antisense direction
WBm0687	Antisense	Predicted protein	Start position elongated 516 nucleotides sense direction
WBm0716	Sense	ABC-type Zn2+ transport system, periplasmic component	Start position elongated 147 nucleotides antisense direction
WBm0722	Sense	ATP-dependent protease HslV, peptidase subunit	Start position elongated 24 nucleotides antisense direction
WBm0730	Sense	Predicted permease	Start position elongated 75 nucleotides antisense direction
WBm0733	Sense	1-acyl-sn-glycerol-3-phosphate acyltransferase	Start position elongated 138 nucleotides antisense direction
WBm0763	Antisense	Asp-tRNAAsn/Glu-tRNAGln amidotransferase A subunit	Start position elongated 33 nucleotides sense direction
WBm0788	Sense	Preprotein translocase subunit SecD	Start position elongated 33 nucleotides antisense direction

Table 2-2: List of all 25 modifications made to existing open reading frames in *wBm*'s genes, based on the presence of RNA-sequence data to these regions and homology from BLAST-search results.

2.2.3 Bioinformatic characterisation of genes of interest

RNA-sequencing provides direct experimental evidence for transcription as compared to inferred annotations and was used to prioritise some of the newly identified genes for further investigation. One in particular (*wBmNew0001*) showed very high expression levels in *wBm* during the studied time-point (>1,000 FPKM). In addition, initial BLAST searches have identified possible homologues present in several *Wolbachia* of supergroup D. This supergroup, along with supergroup C, contains *Wolbachia* that are exclusively endosymbionts of nematodes, with several examples of *Wolbachia* hosts in supergroup D including *B. malayi*, *Wuchereria bancrofti*, and *Litomosoides sigmodontis*.

InterProScan[122] searches of *wBmNew0001* predicted the presence of two transmembrane domains near the N-terminus of the protein and a long, low-complexity tail at the C-terminus (**Figure 2-6**). The positions of the transmembrane domains and the orientation of the protein was confirmed by TMHMM and Phobus[186,199]. Further PSI-BlastP[200,201] searches identified the acyl-CoA dehydrogenase family as having conserved regions that aligned to *wBmNew0001*. Upon further alignment using the t_coffee[202] program, a region of 10 residues was identified, 6 of which were conserved across a number of acyl-CoA dehydrogenases (Thre[152], Glu[153], Gly[155], Asp [159], Ala[160], Ala[161]). Comparison to the RCSB Protein Databank[183] identified this region as partially overlapping with the predicted active residues in an FAD-binding domain (RCSB ID number 1RX0_A[203], Tyr[136], Leu[138], Thre[139], Gly[144], Ser[145], conserved residues underlined) (**Figure 2-6**). This domain is found specifically in Isobutyryl-CoA dehydrogenase, responsible for valine catabolism. However, this partial overlap of two residues makes it unclear if this region remains functional, as this FAD binding site is also predicted to use three other domains to form an active binding pocket.

Although initial searches identified homologues present in only *Wolbachia* of nematodes, further BLAST, PSI-BLAST, and manual alignments were done to identify if any homologues of this gene exist within any other sequenced *Wolbachia* strains. These searches identified single homologues in many *Wolbachia* strains, with no related genes that can be identified outside of the *Wolbachia* genus. Using this data, comparative sequence alignments were then constructed using the program T_Coffee[202] and a selection of these sequences. From this, we identified 3 regions of the gene that showed high levels of conservation, as well as a large low-complexity C terminus region (**Figure 2-6**). Interestingly, it was observed that the extent of similarity in this C terminus region appeared to be associated with *Wolbachia* supergroup.

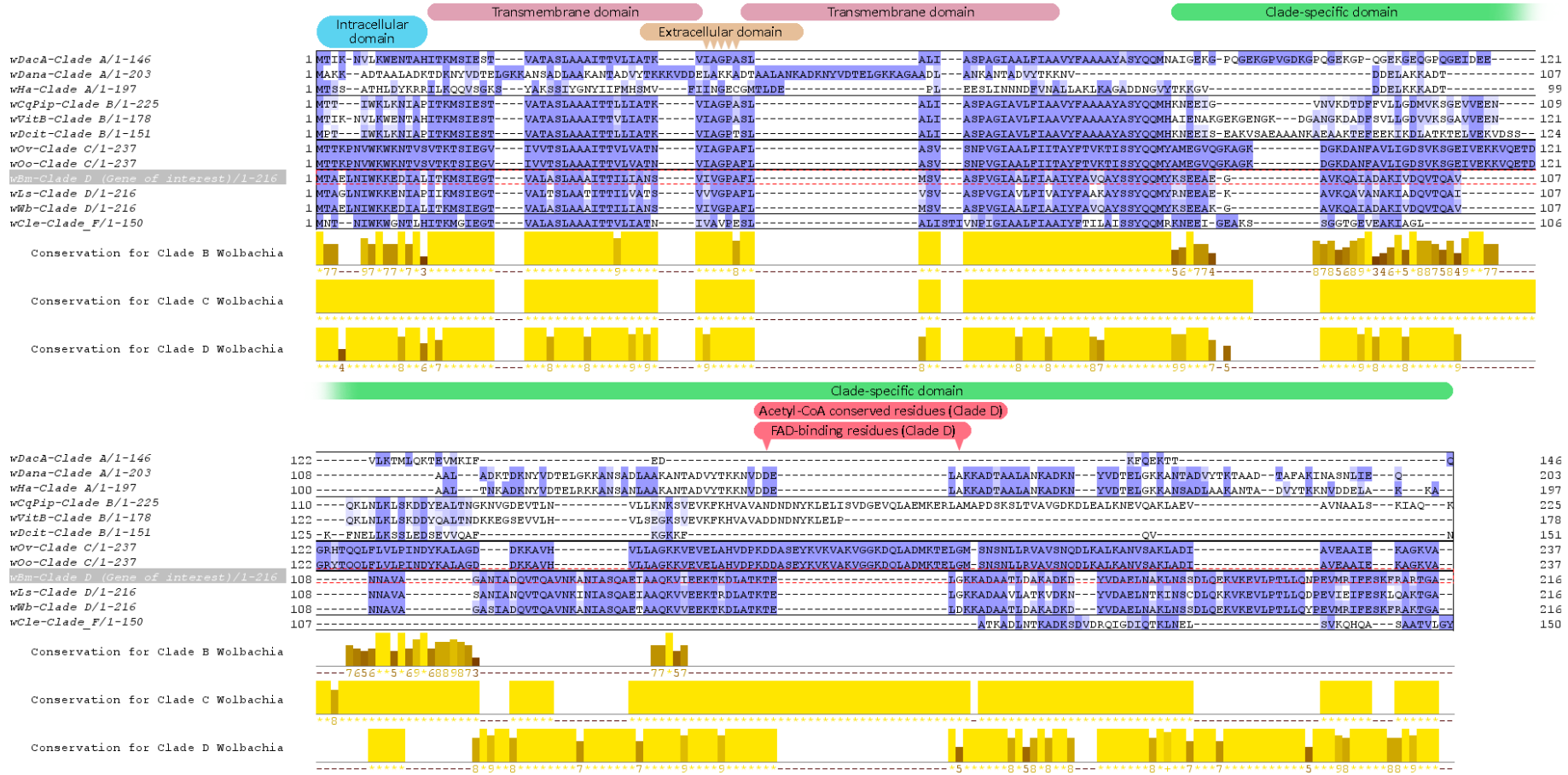


Figure 2-6: Comparison of the newly annotated gene wBmNew0001 to homologous genes found in a small sample of other* Wolbachia *strains, organised by* Wolbachia *supergroup. *The different coloured regions of the alignment correspond to their predicted domains. Regions in **green** correspond to transmembrane domains. Regions in **pink** correspond to extracellular domains. Regions in **purple** and **blue** correspond to intracellular domains, with regions in **purple** being highly conserved across all compared supergroups, and regions in **blue** being highly variable, with this variation possibly conserved by different supergroup. Note the yellow histograms below the alignments, showing high levels of conservation within this sequence between the different supergroups, specifically the two transmembrane regions and the domains surrounding it.*

2.2.4 Overview of newly identified pseudogenes

In addition to new protein coding genes, this study was able to identify 104 new potential pseudogenes (**Figure 2-4**, green and blue segments), in addition to the 110 identified in 2005. These 104 new potential pesudogenes excludes 3 newly identified functional RNAs and 5 instances of non-model translational events. Analysis of the new pseudogenes identified 29 with homologues to genes of known function in existing databases, with the remaining 74 having no database matches or matches to genes of unknown function (**A 1-4**). 25 of these regions showed high FPKMs (>100) despite their size (approximately 200 nucleotides) and lack of annotations in existing databases.

The rate of pseudogene occurrence has been observed to vary significantly between *Wolbachia*, ranging from 7 to 21% of the annotated genome[118]. The combined presence of 214 pseudogenes represents approximately 20% of *w*Bm's total gene count.

2.2.5 Potential Non-Standard Translational Events in *w*Bm

The central dogma of molecular biology describes how genetic information in the form of DNA is processed into mRNA via transcription by RNA polymerases, before being synthesised into proteins via translation by ribosomes. Traditionally, it was expected that translation only occurred on a single reading frame of DNA, and started or stopped upon the ribosome's encounter with a corresponding start or stop codon. In the past, it has been known that viruses often do not follow these rules, and are able to read past stop codons, or even shift reading frames part-way through translation, thus producing proteins that may have different functions[204–206]. Recently however, this phenomenon has been observed occurring in a range of prokaryotes[207] and eukaryotes[208,209], thus generating new levels of complexity within certain organisms.

Specifically in *w*Bm, potential sites of such 'non-standard translational processes' were identified during the course of reannotating *wBm*'s genome. These included 2 instances of Stop-Codon Read-through (SCR, *wBmNew0101*, *wBmNew0102*), which are evidenced by the presence of aligned RNA-sequence reads to a large open reading frame interrupted by at least one stop-codon, and high identity to homologues in other sequenced *Wolbachia* genomes. An additional 3 instances of Programmed Ribosomal Frame-shifting (PRF, *wBmNew0103* to *wBmNew0105*) were also identified based on the presence of aligned RNA-sequence reads spread across 2 different reading frames, as well as high identity to homologues in other sequenced genomes of *Wolbachia* or other organisms (typically >90% sequence coverage for the full product). Three cases of PRF resulting in reclassification and extension of existing genes from 2005 (**Figure 2-7**).

All 5 of the mentioned regions had significant expression quantities across the length of the transcript, while BLAST searches across the full length of these regions revealed nearly 100% identity homologues in other *Wolbachia* with known function. This is despite the stop-codon interruptions or frame-shifts the genes have undergone. Taken together, these 4 points of evidence has led to the 5 genes in question being classified as cases of SCR or PRF. These regions were possibly missed by Foster *et al.* in 2005 due to developments in knowledge, and the lack of RNA-sequence technology to allow for confirmation of gene expression of these regions.

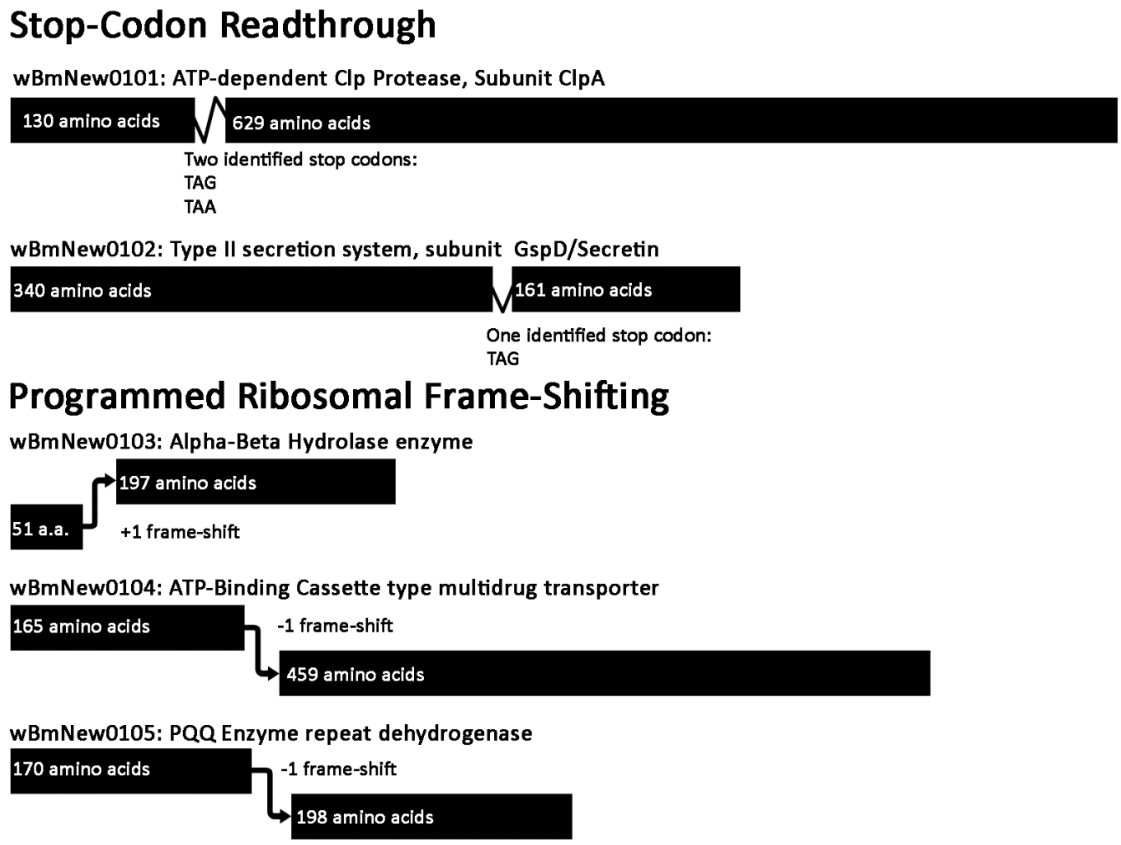


Figure 2-7: Overview of the 5 genes that show evidence of either Stop Codon Read-through or Programmed Ribosomal Frame-Shifting.

An example of SCR includes a protein homologous to the ATP-binding subunit of a Clp protease, or the ClpA subunit (*wBmNew0101*, **Figure 2-7**), which was originally identified by Foster *et al.* in 2005 as a pseudogene. This protein has been interrupted by two stop codons, with the average FPKM of these two fragments indicating high expression (>150 FPKM), and thus a potential case of SCR. This region is homologous to Clp proteins that can be found in other *Wolbachia*. Analysis of *wBm*'s existing genome indicates the presence of 3 other Clp protease subunits (ClpB, ClpX, ClpP), all of which are highly expressed (>100 FPKM). While

the ClpP subunit has some proteolytic capability on its own, it requires ClpA to obtain effective protease activity for regulatory protein degradation[210–213]. Separately, the ClpB and ClpX subunits can also bind to the ClpP subunit, which confers either protein fold-refolding capability, or DNA damage repair and protein quality control activity respectively[213–215].

An example for PRF can be found in *wBmNew0104* (**Figure 2-7**), which encodes for the permease component of an ATP Binding Cassette (ABC) Transporter system, separated over two reading frames. Despite this interruption, the full gene has up to 94% query coverage and 84% identity with variants of ABC permeases found in other sequenced *Wolbachia* and other members of the alphaproteobacteria class, such as *Orientia tsutugamushi* or *Caedimonas varicaedens*. Its high expression level (150 FPKM) is also indicative of active transcription. Analysis of *w*Bm's existing genome indicates it maintains many similar ABC transporters that work to move a diverse range of metabolites (such as haem or phospholipids) and metal ions (such as iron (III) or zinc) across the bacterial membrane. While BLAST searches of *wBmNew0104* identified its function as an ABC transporter, no homologues with a known substrate were found that could indicate the specific substrate for this gene.

2.2.6 Comparison of *w*Bm genes against other *Wolbachia*

As the genome of *w*Bm was the first of many *Wolbachia* genomes to be sequenced, it serves as an important template for all succeeding *Wolbachia* genomes. Thus, it can be assumed that these 24 newly identified genes may have been missed in these newly sequenced genomes. As such, we then attempted to investigate if any of these 24 genes are present in these sequenced *Wolbachia* genomes, and if any may be conserved only in mutualistic strains, and not in strains that present a parasitic phenotype. For this, a list of 30 different *Wolbachia* genomes encompassing strains from supergroups A to E, F and L were downloaded from NCBI and nematode-specific databases (**Table 2-3**). This list of 30 *Wolbachia* strains was chosen based on their supergroup distribution, spread across 7 nematode and 23 arthropod hosts, and the existence of a publicly available genome (**Table 2-3**). These were compared against the genes of *w*Bm via the Large-Scale Blast-Score Ratio program[216]. This utilises BlastX searches of a list of query proteins against a list of test genomes, resulting in a scoring matrix of the individual genes against the test genome. The given score represents the percentage identity of the query gene in the test genome, ranging from no identity (0%) to full match (100%). For the purpose of this investigation, a homologue is defined as existing in a given genome if it has over 30% identity to the test *w*Bm gene, a threshold that has been referred to as the 'twilight zone' of bioinformatics[217].

Of the 24 newly identified genes, there is a core set of 8 genes that have homologues in 26 of the 30 analysed genomes, with a further 2 genes that have homologues in 19 genomes and a final 3 genes that have homologues in 12 genomes (**Figure 2-8**). An additional 7 genes were identified as unique to *Wolbachia* of *Wuchereria bancrofti* (*w*Wb) and *w*Bm, with only 1 exception (*wBmNew0010* in *Wolbachia* of *Folsoma candida,* **Figure 2-8**).

The genome of *w*Wb is known to be closely related to *w*Bm[218], thus the identification of homologues for all 24 genes was not surprising. This was followed by the *Wolbachia* endosymbiont of *Litomosoides sigmondontis* (*w*Ls), a filarial parasite of rodents. *w*Ls is a member of supergroup D like *w*Wb and *w*Bm, which showed 15 of the 24 newly identified genes having homologues. Conversely, the genome with fewest homologues, at only 7, is the genome of *Wolbachia* of *Pratelynchus penetrans* (*w*Ppe), a plant parasitic nematode (**Figure 2-8**). Between these two extremes are a mix of the 23 insect-infecting *Wolbachia*, and the 3 nematode-infecting *Wolbachia* of Supergroup C (*Wolbachia* of *Dirofilaria immitis*, *Onchocerca ochengi* and *Onchocerca volvulus*).

Expanding this analysis to the entirety of *w*Bm's known genome, a similar picture emerges with a core set of genes that are highly conserved amongst all the *Wolbachia* strains (**Figure 2-9**). Again, the genomes of *w*Wb and *w*Ls again appear to be most closely associated with the *wBm* genome, whilst *w*Ppe shows the lowest level of conservation. Interestingly, the three *Wolbachia* strains of filarial nematodes from Supergroup C (*Wolbachia* of *Dirofilaria immitis*, *Onchocerca volvulus*, and *Onchocerca ochengi*) do not show any unique blocks of conserved genes as compared to the remaining 23 arthropod-infecting *Wolbachia* (**Figure 2-9**). Indeed, the genes of *w*Bm appear to be most specialised, or divergent, amongst *Wolbachia* of other filarial nematodes when compared to the genes of arthropod-host *Wolbachia*. Amongst nematode *Wolbachia* outside of Supergroup D *Wolbachia*, only about 74% of their total gene count showed homology to the *w*Bm genome (**Table 2-3**). The only genome with a lower number of homologues is the genome of *Wolbachia* of *Drosophila ananassae*, and *Wolbachia* of *Muscidufax uniraptor* (Supergroups A and L respectively).

Wolbachia of *M. uniraptor* appears to be an exception to the general rule however, with many *Wolbachia* of insects noted to have high numbers of homologues to *wBm*'s genes. The *Wolbachia* with the highest number of homologues include *Wolbachia* of *C. lectularius* (common bedbug, Supergroup F *Wolbachia*, 85% genes with homology, **Table 2-3**), *Wolbachia* of several *Drosophila* species (Supergroup A, up to 84%), and *Wolbachia* of the *Nomada* genus ('cuckoo bees', Supergroup A, 83%).

Overall, the analysis conducted with LSBSR effectively divides the analysed *Wolbachia* genomes between Supergroup D, and all others, regardless if only the 24 newly identified genes were used, or the whole *w*Bm genome.. Specifically, genomes of *Wolbachia* from Supergroup D frequently scored highly, whilst genomes from other *Wolbachia* supergroups had generally similar scores, with the exception of *w*Ppe which consistently scored the lowest.

Genome ID (NCBI reference ID) or host website	Strain acronym	Strain host	Supergroup	Release Date	Genes > 30% identity (new genes)	Genes > 30% identity (wBm genome)	% of New (% of Genome)
NZ_LSYX00000000.1	*w*DacA	*Dactylopus coccus*	A	2016	11	696	46% (80%)
NZ_AAGB00000000.1	*w*Dana	*Drosophila ananassae*	A	2005	10	613	42% (71%)
AE017196.1	*w*Mel	*Drosophila melanogaster*	A	2004	13	721	54% (83%)
NZ_AQQE00000000.1	*w*MelPop	*Drosophila melanogaster*	A	2013	13	720	54% (83%)
CP003884.1	*w*Ha	*Drosophila simulans*	A	2013	14	725	58% (84%)
NZ_CAOU00000000.2	*w*DSuzu	*Drosophila suzukii*	A	2012	14	725	58% (84%)
NZ_AWUH00000000.1	*w*Gmm	*Glossina morsitans morsitans*	A	2014	11	669	46% (77%)
NZ_ACFP00000000.1	*w*Uni	*Muscidufax uniraptor*	A	2017	11	589	46% (68%)
NZ_MUJM00000000.1	*w*VitA	*Nasonia vitripennis*	A	2017	13	712	54% (82%)
NZ_LYUY00000000.1	*w*Nfe	*Nomada ferruginata*	A	2016	14	717	58% (83%)
NZ_LYUW00000000.1	*w*Nfla	*Nomada flava*	A	2016	14	717	58% (83%)
NZ_LYUV00000000.1	*w*Nleu	*Nomada leucophthalma*	A	2016	14	717	58% (83%)
NZ_LYUX00000000.1	*w*Npa	*Nomada panzeri*	A	2016	14	717	58% (83%)
NZ_CAGB00000000.1	*w*AlbB	*Aedes albopictus*	B	2018	11	700	46% (81%)
NZ_CTEH00000000.1	*w*Pip.Mol	*Culex pipens molestus*	B	2013	11	709	46% (82%)
AM999887.1	*w*CqPip	*Culex quinquefasciatus JHB)*	B	2008	11	711	46% (82%)
NZ_LSYY00000000.1	*w*DacB	*Dactylopus coccus*	B	2016	10	690	42% (80%)
NZ_AMZJ00000000.1	*w*Dcitri	*Diaphorina citri*	B	2013	10	704	42% (81%)
CP003883.1	*w*No	*Drosophila simulans (Noumea)*	B	2013	11	707	46% (82%)
NZ_AERW00000000.1	*w*VitB	*Nasonia vitripennis*	B	2011	10	696	42% (80%)
NZ_JYPC00000000.1	*w*BaOb	*Operophtera brumata*	B	2015	10	698	42% (81%)
HTTP://NEMATODES.ORG/GENOMES	*w*Dim	*Dirofilaria immitis*	C	2012	9	637	38% (74%)
HE660029.1	*w*Oo	*Onchocerca ochengi*	C	2012	8	640	33% (74%)
HG810405.1	*w*Ov	*Onchocerca volvulus*	C	2013	8	640	33% (74%)

NC_006833.1	*w*Bm	*Brugia malayi*	D	2005	24	866	100% (100%)
HTTP://NEMATODES.ORG/GENOMES	*w*Ls	*Litomosoides sigmondontis*	D	2012	14	724	58% (84%)
NZ_NJBR00000000.2	*w*Wb	*Wuchereria bancrofti*	D	2017	24	855	100% (99%)
NZ_CP015510.2	*w*Fol	*Folsoma candida*	E	2017	11	667	46% (77%)
AP013028.1	*w*Cle	*Cimex lectularius*	F	2014	12	731	50% (84%)
NZ_MJMG00000000.1	*w*Ppe	*Pratylenchus penetrans*	L	2016	7	380	29% (44%)

Table 2-3: List of* Wolbachia *genomes used in the analysis, sorted by supergroup, listing out ID numbers, host organism, and sequencing dates.

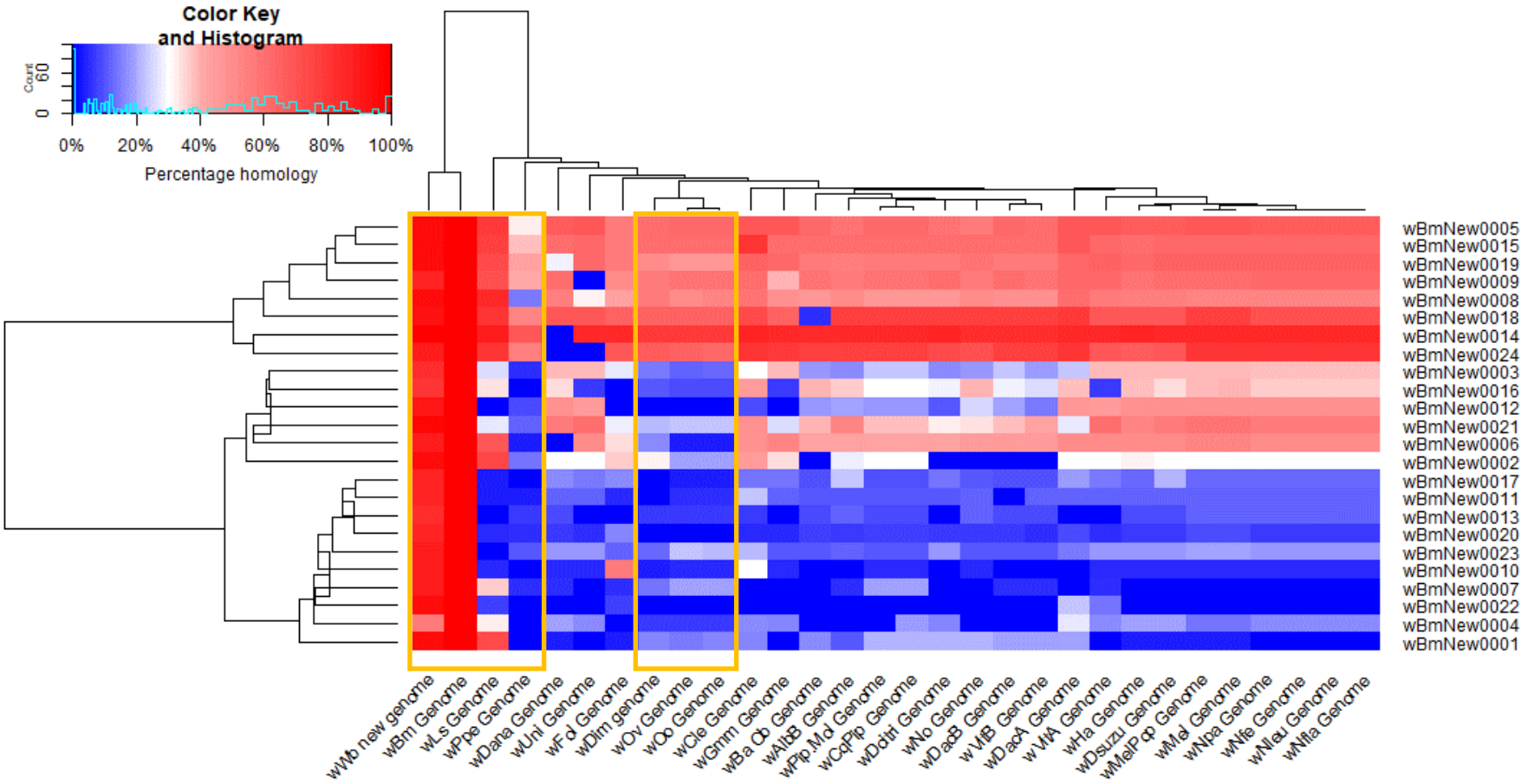


Figure 2-8: Heatmap to show levels of conservation amongst the 24 newly identified genes from this study, amongst a selection of 30* Wolbachia *genomes. *Red indicates that the specific* w*Bm gene (row) has a homologue with high identity in the specific* Wolbachia *genome (column), whilst blue indicates that the gene has no homologue within the specific* Wolbachia *genome. Note the yellow highlighted columns which indicate all nematode-*Wolbachia *endosymbiont genomes.*

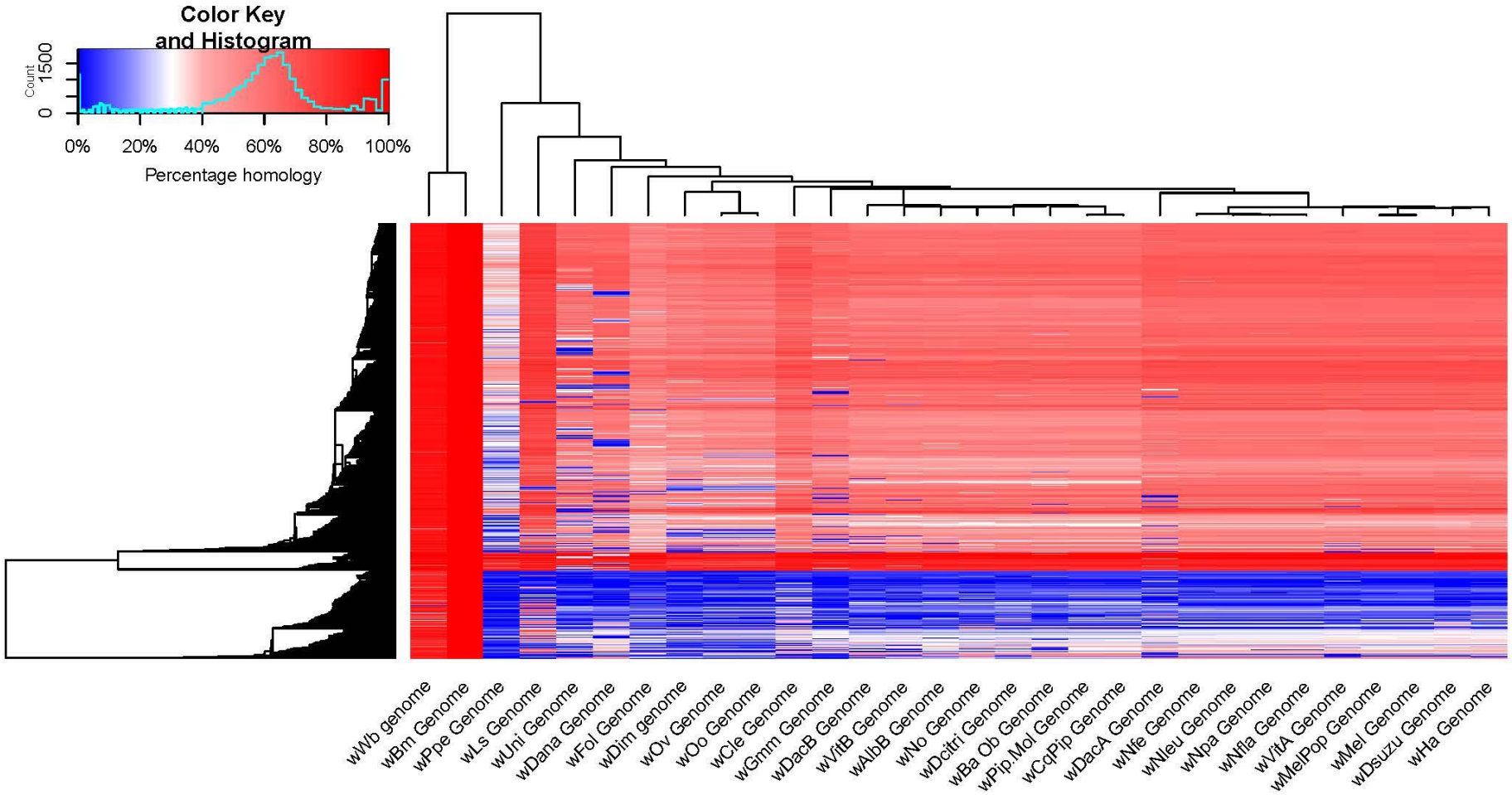


***Figure 2-9: Heatmap to show levels of conservation of all coding* wBm *genes, including those identified in this study, rRNAs and tRNAs, amongst a selection of 30* Wolbachia *genomes*.**

2.2.7 Survey of potential of Single Nucleotide Variants in the *w*Bm genome

RNA-sequencing data provides an opportunity to check for the presence of Single-Nucleotide Variants (SNVs) in coding regions by comparing expressed transcripts to the *wBm* genome. This information is useful to validate the 24 newly identified genes, as well as to check for the presence of sequencing errors or variant mutations. Whilst there are a variety of programs that can been used to identify variants, many of these are primarily designed for population genomics studies rather than variant calling with RNA-sequence data. These programs also suffer from either high sensitivity and low positive-predictive value (PPV), or vice-versa, and often have minimal concordance with each other[219,220]. As such, we utilised SAMTools/BCFTools[221] for variant identification, chosen for its predicted balance between sensitivity and PPV, and workflows that make it suitable for applying to RNA sequence data. After quality filtering these variants (using vcfutils by SAMTools[221]), the program SnpEff[222] was utilised to summarise and predict the severity of impact by analysing the location of these variants and, if located within protein-coding genes, their predicted effects on the protein coding sequence. Such effects could include a reading-frame shift due to an insertion or deletion (which are considered to have high impact), to non-synonymous single nucleotide polymorphisms (SNPs) that could cause non-synonymous mutations in protein coding sequence (moderate impact), or synonymous SNPs in protein coding sequence (low impact). The presence of such variants were checked against both the 24 newly identified genes, as well as the genome of *wBm* in general via the newly-generated GFF files from this study.

Initial variant calling and filtering identified 530 variants within the *w*Bm transcriptome, with the majority of these variants being classified as deletion mutations (446 deletions, **Table 2-4**). Of the 530 identified variants, 328 were predicted to have a high impact on *w*Bm genes, 23 predicted to have moderate impact, and 41 predicted to have low impact (**Table 2-4**). However, upon manual inspection of variant call output files, the majority of deletions identified by SAMTools are due to deletions of single nucleotides in low complexity regions, within strings of 2 or more of the same nucleotides (69.5%). Subsequent visual inspection using the Integrated Genome Viewer (IGV) program[223] further identified one of the replicates as appearing to have a significantly higher number of SNPs as compared to the other replicates (**Figure 2-10**), again the majority of these being deletions. Upon excluding this replicate from the analysis, the number of variants identified was reduced significantly, with none of these variants being due to deletion mutations and the majority being due to SNPs (55 identified variants, 49 due to SNPs, **Table 2-4**). These observations imply errors within the machine used to conduct the sequencing for the one replicate, leading to difficulties in

accurately calling strings of the same nucleotide. As sequencing inaccuracies would impact the ability for accurate variant calling, replicates from sample 2 were ignored for this analysis.

Variant type	BCFTools (all replicates)	BCFTools (no replicate 2)
SNP	79	49
Insertion	5	6
Deletion	446	0
Total	530	55
High impact	328	5
Moderate impact	23	27
Low impact	41	13

Table 2-4: Summary of variants, and their predicted impacts, identified in the transcriptome of* wBm *when compared to the 2005 sequenced genome. *Note the significantly reduced number of deletion variants once the second replicates were removed from analysis, and the resulting effects on variants deemed as having a 'high impact' on genes in* wBm*.*

The number of identified variants in the *w*Bm transcriptome is higher than the 18 experimentally confirmed variants identified in the recent *w*Bm genome resequencing efforts[138], although these 18 are solely experimentally validated variants. Of these, only one is predicted to have any direct impact on the 24 newly-identified genes. This affects the 216-residue long gene *wBmNew0001*, altering residue 211 from an Arginine to a premature stop codon (CGA → TGA), thus being classified as a 'high' impact variant.

Outside of the 24 newly identified genes, 5 variants were predicted to have high impacts on the previously identified 805 genes of *w*Bm. These affect *wBm0339* (large ribosomal subunit protein L23), *wBm0458* (F-type ATP synthase subunit b), *wBm0540* (aconitate hydratase), *wBm0692* (ribonuclease P), and *wBm0738* (an acyl carrier protein). All variants were due to insertions of additional adenine or thiamine nucleotides within their coding sequences, thus causing a frame-shift mutation. It is interesting to note that these insertions are frequently part of a long string of similar nucleotides, indicating possible sequencing errors. The full list and details of identified SNVs is included in **Appendix 1** as a Variant Call File (VCF, **A 1-6**).

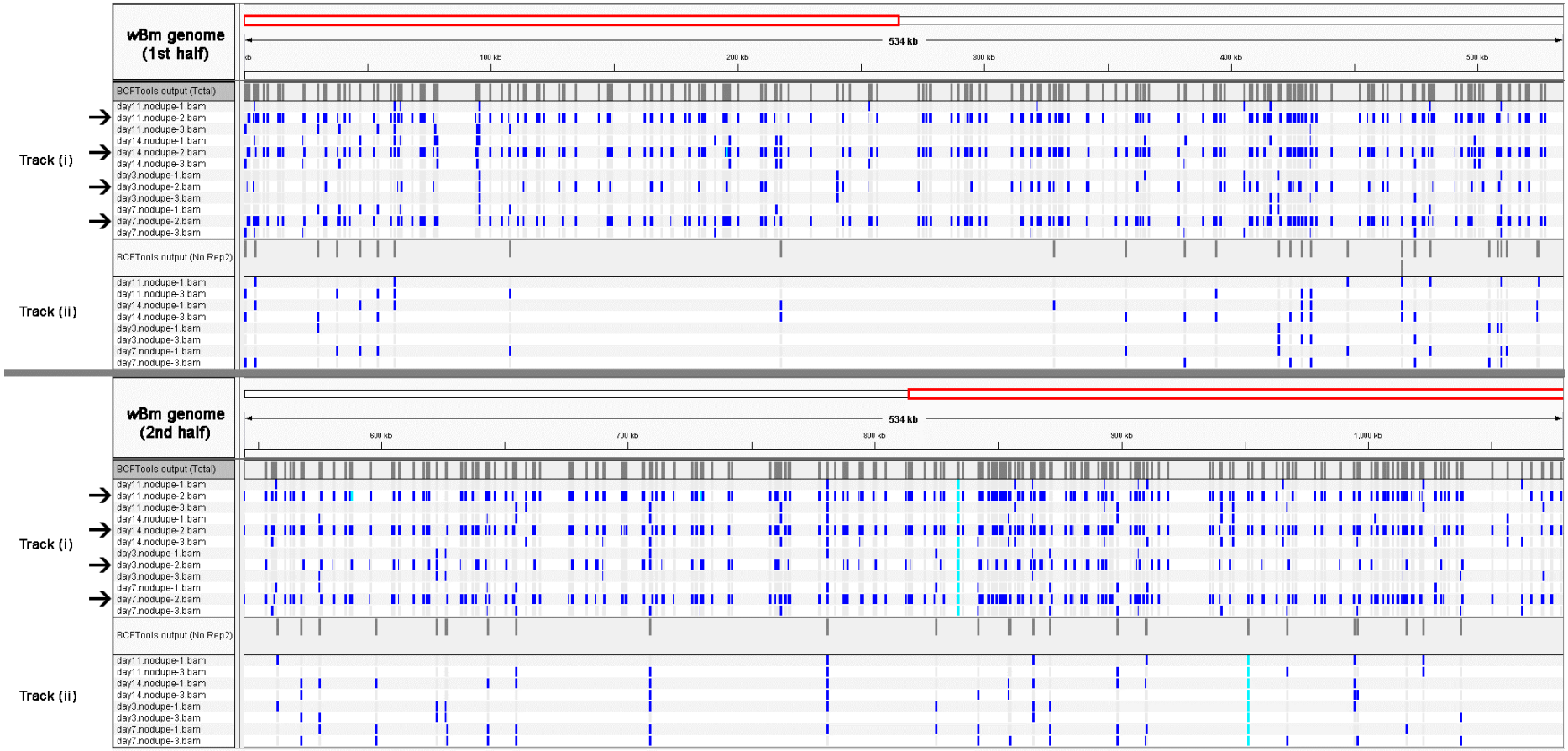


Figure 2-10: Visualisation of wBm's entire genome split over two lines, and the variants identified by BCFTools, separated based on replicate samples used for analysis. *Each blue bar in the image identifies a single variant within the test sample/replicate, with the bar colour varying based on call quality. Note the two separate tracks;* ***Track (i): BCFTools output (Total)*** *and* ***Track (ii): BCFTools output (No Rep2)****. Note that in* ***Track (i)****, there are test samples which contribute to the majority of variants identified (arrows), all of which stem from a particular technical replicate. Upon removing these replicates in* ***Track (ii)****, these identified variants are removed.*

2.3 Discussion

The newly identified genes from this study highlights the importance of reannotating genomes by incorporating new experimental data. Here we report and describe a total of 24 new coding genes, 3 of which encode functional-RNAs, an additional 104 potential pseudogenes, as well as 5 cases of non-standard translational events. These non-standard translational events are unlikely to have been identified based on simple automated algorithms, which would have classified them as pseudogenes. Knowledge of these pseudogenes is particularly important when dealing with a bacterial genus as diverse as *Wolbachia*, aiding phylogenetic studies and identifying cases of genes that may have been lost or gained over the course of evolution and possibly identifying genes that play a role in symbiosis. We also report changes to 25 ORFs in the *wBm* genome, identified with the aid of RNA-sequence alignments and BLAST results against existing databases (**Table 2-2**).

During the course of the reannotation and validation work, NCBI independently released an updated version of the *w*Bm genome that had been processed by the Prokaryotic annotation Pipeline[120] (accession ID GCF_000008385.1_ASM838v1). This annotation identified a total of 839 protein-coding genes and 167 pseudogenes. Interestingly, this automated annotation method had reclassified several of Foster *et al*.'s protein coding genes as pseudogenes or splitting individual genes into two separate pseudogenes. Upon manual comparison between the two annotations (from this study and NCBI's newly released annotation), it was noted that all 24 genes identified from this study were also identified in NCBI's automated annotation. Furthermore, whilst NCBI's study identified an additional 63 genes when compared to this study (both functional RNAs and protein-coding genes), 55 of these genes were identified as being hypothetical with no known function, and only 2 were noted to code for proteins with known function. As to why these were not identified during this study, the vast majority of these genes were noted to be small (<300 nucleotides in length) that did not map well with RNA-sequence data generated from this study, the driving force behind this reannotation work. While these factors by themselves do not indicate the presence of spurious annotations, this study specifically sought to avoid such false-positives, and their subsequent propagation in databases, via use of RNA-sequencing, gene size cut-offs, and presence of homologues as validation metrics. Nevertheless, as all new, independently identified genes are shared between the two annotations, this highlights the accuracy of an RNA-seq based approach to the reannotation of genomes.

Furthermore, the SecE gene has also been identified in a previous study by Frutos *et al.*[224]. This gene was identified via manual investigation of the *w*Bm genome as part of comparative analysis on genomes, gene synteny and organisation in the bacterium *Ehrlichia ruminantum*, Gardel and Welgevonden strains, and other members of the order Rickettsiales. This bacterium is an alpha-proteobacterium, a distant relative of *Wolbachia pipientis*, and the causative agent of heartwater, a tick-borne disease that affects ruminant livestock in Africa, and has been recently introduced into the Americas. Frutos *et al.*[224] notes that this gene was likely missed by Foster *et al.* due to the small size of this gene, at only 69 amino acids. Again, this emphasises the benefits of applying RNA-sequencing data to better identify such small genes.

The annotation of SecE in the *w*Bm genome is important for the understanding of the symbiotic relationship, as are 4 other newly-identified genes in *wBm* which encode components for both Type IV and Sec-Secretion transport mechanism components (**Table 2-1**). This SecE subunit functions as a key component of the Sec-translocase system for either protein excretion or embedding into the bacterial membrane. Whilst the Type IV secretion system was previously demonstrated to exist as a functional unit in *w*Bm using microscopy and proteomics experiments[159,225], the original genome sequence of *w*Bm lacked two key components of the pathway, the VirB2 and VirB7 subunits, which were identified as part of this study. Both of these systems likely serve vital roles for obligate endosymbionts like *w*Bm: The Type IV secretion systems for instance are broadly distributed amongst bacterial species[194,226], and are well-known effector-translocator systems for directly transporting molecules or DNA into eukaryotic/host cells, or between other bacteria/*Wolbachia*. For instance, the distantly related *Rickettsia* genus is comprised of obligate intracellular endosymbiotic bacteria, with all known members encoding a related Type IV secretion system[140]. The Sec secretion system meanwhile plays a role in insertion of proteins to the extracellular bacterial membrane. This is an important process in *w*Bm, as it is known to maintain a host of surface proteins that can interact with both the nematode or mammalian host for processes such as motility within the nematode, or eliciting pathology and aiding nematode immune system evasion within the mammalian host[88,89,152,227].

This study also identified three functional-RNA coding genes (**Table 2-1**) by comparison to the RFAM database[184]. These RNA coding genes were noted to have extremely high FPKMs (>1,000 FPKM), and have been identified as being ubiquitous in other bacteria. They fulfil critical roles in tRNA maturation[228] (bacterial RNAse class P, *wBmNew0015*), transcription regulation[229,230] (6S RNA, *wBmNew0010*), as well as the rescuing of stalled ribosomes due to

incorrectly-sequenced mRNAs[231,232] (alpha transfer-messenger RNA, *wBmNew0019*)[228–232]. All of these features appear to be widespread in bacteria and allow maintenance and basic survival. The lack of annotation for these functional RNAs by Foster *et al.* in 2005[111] is likely due to these genes being highly interrupted with stop-codons, as well as a lack of RNA-sequencing technology , making it difficult to identify these small genes.

Of the 24 newly identified genes, 8 genes were noted to have homologues in databases that have no known function or did not have any known homologues at all. Prioritisation of one gene (*wBmNew0001*) due to its expression level, and subsequent characterisation by InterProScan[122] and BLAST searches against the Protein Data Bank[183], identified 2 transmembrane regions, indicating its localisation within a membrane layer. Subsequent BLAST searches and alignments against other organisms identified homologues only in other *Wolbachia*, with a clear association between domain and *Wolbachia* supergroup. Whilst the transmembrane regions remain perfectly conserved amongst all homologues regardless of supergroup origin, the protein tail's variability appears conserved by supergroup, pointing to an evolutionary relationship (**Figure 2-6**). This tail has been predicted by multiple programs to lie within the cytoplasm of *Wolbachia*, making it unlikely to have a role in directly interacting with the nematode host. However, such low-complexity regions are increasingly being associated with protein-protein interactions within the literature[233], and may instead act as a form of excretory/secretory mechanism for *Wolbachia*, specialised for the specific supergroup relationship via formation of a transmembrane pore as a multimeric complex. For instance, while it was found that a partial flavin adenine dinucleotide (FAD) binding site was present within this cytoplasmic tail of *wBm* and other Supergroup D *Wolbachia*, this binding site is absent within *Wolbachia* from other supergroups (**Figure 2-6**). At least amongst *Wolbachia* of Supergroup D and C (*Wolbachia* of nematodes), *Wolbachia* of Supergroup D are the only known group that can produce FAD *de-novo*[142], whilst their associated nematode hosts are unable to do so and require *Wolbachia* to supply this key cofactor.

Outside of these protein- or functional RNA- coding genes, this study was able to identify 2 cases of Stop Codon Readthrough (SCR) and 3 cases of Programmed Ribosomal Frame-shifting (PRF) (**Figure 2-7**). Such translational processes are most commonly known in viruses, and have been known to exist in bacteria for many years[205,234] (also reviewed by Baranov *et al.*[235] and Atkins *et al.*[236]), although such events are not as well-studied in eukaryotes. Both processes are believed to play a role in gene regulatory mechanisms[235,236], particularly in response to external stressors, possibly in the form of environmental changes or introduction

of chemicals like antibiotics. The occurrence of PRF events appear to be governed by the presence of 'shift' sites, which are encoded into the coding nucleotide sequence itself, and their interactions with bacterial Release Factors 1 and 2[236] (of which *w*Bm maintains both). These release factors are responsible for recognising stop-codons in mRNA sequences and forces the ribosome to stop translation of peptides once these are recognised. Due to the ability of these release factors recognising only 2 of the 3 different potential stop-codons each (TAA or TAG in release factor 1, and TAA or TGA in release factor 2)[237,238], the same release factors also play a role in permitting SCR by ribosomes, with these stop codons being replaced by an alternative amino acid[239].

Specifically in *Wolbachia*, none of these non-standard translational events have yet been recorded, and may be worth additional study. If the mechanisms that allow translation of these genes into functional proteins do exist in *wBm*, this points to a more complex ability for *wBm*, and possibly other *Wolbachia*, to regulate protein production across their diverse host range. Specifically, these 5 instances of non-model translational processes potentially encode proteins that have proteolytic or secretory functions (**Figure 2-7**). Such genes in *w*Bm may be selectively expressed in response to environmental changes, such as those encountered during the complex, lifecycle of its nematode host.

This study was also successful in identifying 104 potentially new pseudogenes, several of which had high FPKM values. One potential reason for this could be that the aligned RNA-sequences transcribed from them may be novel regulatory RNA sequences[240,241]. Several of these potential regulatory RNA sequences are found neighbouring other large genes or operon regions.

Comparing the 24 newly identified genes to the genomes of 30 additional *Wolbachia* species identifies approximately 3 different clusters of genes that show differing levels of identity (**Table 2-3**, **Figure 2-8**). Analysis via the Large-Scale Blast Score Ratio system was initially performed to check if any of the 24 newly-identified genes may be present in only mutualistic *Wolbachia*, such as those found in nematodes, or perhaps even *Wolbachia* of *Cimex lectularius* (*w*Cle, common bedbug). However, no such relationship appears to exist. When the average percentage of genes that show homology to *w*Bm amongst supergroups are compared, it becomes clear that Supergroup C, which consists of solely mutualistic *Wolbachia*, actually has a lower percentage of genes with evidence of homologues to *w*Bm as compared to Supergroups A and B, which consists of *Wolbachia* best-known for reproductive parasitism (**Table 2-3**, averaged 34.67% versus 52.85% and 43.70%

respectively). The *Wolbachia* supergroup with lowest percentage identity is Supergroup L, which for this study contains only *Wolbachia* of *Pratylenchus penetrans*, a plant-parasitic nematode[65]. The low percentage identity for this strain of *Wolbachia* as compared to *wBm*'s new genes may be due to a distant evolutionary relationship, as it has been predicted that plant-parasitic nematodes may be the original host of *Wolbachia*[65].

Broadening this analysis to all previously identified genes of *wBm*, a similar trend can be observed to that of the 24 newly identified genes. Specifically, that there are no distinct groups of genes conserved solely between mutualistic *Wolbachia*, such as those found in Supergroups C and D. Instead, there are 3 distinct groups of genes: One conserved only within *w*Bm and *w*Wb, a second group that is broadly conserved to varying degrees across all *Wolbachia*, and a final group of genes that are highly conserved across all *Wolbachia*, with this group likely encoding for ribosomal genes such as tRNAs or rRNAs (**Figure 2-9**).

Of additional interest is the comparison of the genome of *wCle*, from the common bedbug. The relationship between the two has been shown in the literature to be facultative, based on vitamin B supplementation by the endosymbiont to the host[54,55]. To summarise, *C. lectularius* larvae are unable to develop and moult through the instar stages without *Wolbachia*'s presence, or exogenous Vitamin B supplementation[55]. Genomic comparisons of the vitamin B biosynthetic pathway by Nikoh *et al.*[54] identified the pathway to possibly originate from a lateral gene transfer event by a common ancestor of *w*Cle and *Wolbachia* of the parasitic nematode *Onchocerca ochengi* (*w*Oo), with the operon being later lost in *w*Oo. Nikoh *et al.*[54] were also only able to identify a significantly eroded, and non-functional *de-novo* vitamin B biosynthetic pathway within *w*Bm.

Overall, this genomic comparison of *wBm* to other *Wolbachia* genomes indicates that there is a 'core' set of genes that all *Wolbachia* maintain homologues of, with additional genes that appear to be strain specific. This is evident in the blocks of conserved genes between *w*Wb and *w*Bm with many other *Wolbachia* genomes (**Figure 2-9**). Outside of this core set, there appear to be no distinct groups of genes that contribute directly to a mutualistic phenotype. Such observations suggest that the parasitic phenotype which typically characterises *Wolbachia* of insects is an acquired trait by specific *Wolbachia* strains to increase their biological fitness. A causative example of such traits would be the recently identified *Wolbachia* operon responsible for cytoplasmic incompatibility[148], the most well-known form of *Wolbachia*-based reproductive parasitism.

The transcriptomic data obtained from this study can be exploited to identify the presence of Single Nucleotide Variants (SNVs) within the *w*Bm genome. Analysis by BCFTools identifies a total of 55 possible SNVs, only 1 of which is predicted to have a high impact on 1 of the 24 newly identified genes from this study (*wBmNew0001*). However, this premature stop-codon is noted to occur close to the end of the open reading frame (residue 211 out of 216), in a region of low complexity where no distinct conserved domains can be identified (**Figure 2-6**). This suggests that even though the potential impact of such substitution is predicted to be high, its effective impact may not be, meaning that no high-impact SNVs can be identified within the 24 newly identified genes.

Accurate identification of different *Wolbachia* strains' genes is an important process, illustrated by the range of complex phenotypes the genus displays despite their significantly reduced gene density and content. The relative gain and loss of gene function in one strain may have evolutionary significance for how specific phenotypes may have developed in other strains. There is also the possibility that an as-yet unknown evolutionary benefit may exist for maintaining pseudogenised genes[242] to the level that they are still recognisable by homology studies, and indeed, still recognisable by *Wolbachia*'s transcriptional mechanisms. The results described in this chapter provide an important resource for the further characterisation of the *Wolbachia* genus that could be combined with homology analysis with other related species.

2.4 Materials & Methods

2.4.1 Parasites

Brugia malayi larvae were collected from intra-peritoneal infections of Mongolian jirds (*Meriones unguiculatus*) after 3, 7 (2,000 L3 larvae for both time-points), 11 and 14 (1,000 L4 larvae for both time-points) days post infection (p.i.). The laboratory isolate had been maintained by serial passage in jirds and cats since first derived from a human infection in Malaysia more than 40 years ago, and was also from the same isolate used in the 2005 genome sequencing paper[111]. In order to preserve the RNA, L4 in 5 batches of 150-250 nematodes were collected into 150µl PBS and immediately frozen in liquid nitrogen and shipped on dry ice from TRS laboratories to the Liverpool School of Tropical Medicine. 10 nematode larvae were stored individually for *Wolbachia* load analysis by quantification PCR.

2.4.2 RNA extraction from *Brugia malayi* and *Wolbachia*

RNA was extracted from *B. malayi* and *Wolbachia* using the TRIzol Plus RNA extraction kit (Invitrogen), which used a modified protocol for RNA extraction to that given by the manufacturers.

Briefly, L4 batches were defrosted and simultaneously pelleted by centrifugation (Eppendorf) at 4500rpm at 4°C for 10 minutes. The supernatant was removed and 200µl of TRIzol reagent was added to the samples. All batches of L4 were pooled into a sterile 2ml screw cap tube containing ceramic beads of multiple sizes (1.4 - 2.8mm) (CKMix, Bertin). L4 was homogenized at 6000rpm (Minilys, Bertin Instruments) for 4 x 30 seconds, cooling on dry ice for 30 seconds in between.

The homogenized sample was removed from the beads which were then washed with a further 200µl of TRIzol to maximize sample recovery, before storage at -80°C overnight.

The samples in TRIzol were thawed and subsequently incubated at room temperature for 5 minutes. 200µl molecular grade chloroform (Invitrogen) was added for every 1ml of TRIzol reagent, and mixed by vigorously shaking the tube, followed by incubation at room temperature for three minutes. The sample was then centrifuged at 12,000xg for 15 minutes at 4°C. The upper aqueous phase (approximately 600µl) was removed and transferred to a nuclease-free 1.5ml tube. An equal volume of ice cold 70% ethanol was added to the sample and vortexed for 2 - 3 seconds. The tube was inverted several times to disperse any precipitate generated from the addition of ethanol.

Subsequent binding, washing and elution of RNA from the sample was then conducted in accordance to manufacturer's instructions. Final RNA was eluted in 2 x 50µl molecular grade water (Invitrogen).

2.4.3 Illumina RNA Library Preparation and Sequencing

In order to remove excess large ribosomal RNA from both *Wolbachia* and *B. malayi*, total RNA was treated with Terminator 5'Phosphate-Dependent Exonuclease (Epicentre), in accordance to the manufacturer's instructions.

Illumina GA-IIx cDNA libraries were prepared using the ScriptSeq v2 library preparation kit (Epicentre). Two separate paired-end sequencing runs (2x100bp reads) were performed on an Illumina GA-IIx instrument at the Centre for Genomic Research (CGR), University of Liverpool (UK). One additional run using the same chemistry was performed with a similar

instrument at GenePool, University of Edinburgh (UK). Read files for all data were converted from raw basecall to fastq using CASAVA 1.8 (Illumina). The raw Fastq data files were trimmed for the presence of Illumina-specific adapter sequences using Cutadapt (v1.2.1) [243] with the –O 3 option and further trimmed for quality using Sickle (v.1.33) [244]. Quality scores were assessed using FASTQC v0.9.2[245], coded by by Babraham Bioinformatics (http://www.bioinformatics.babraham.ac.uk/projects/fastqc).

2.4.4 Mapping and Quantifying RNA-sequence Data

A strategy to retain RNA-sequence reads via non-unique mapping was utilised, allowing for the preservation of data relating to lateral gene transfer events, which would potentially be lost when mapping to a concatenated genome. RNA-sequence reads from the processed fastq files were aligned to the genome of *B. malayi* via Subread-Aligner (v. 1.5.0)[188], giving a Binary Alignment/Map (BAM) output file that contained aligned and unaligned sequences. The BAM file was separated into aligned and unaligned reads using Samtools (v. 0.1.19)[221], with one BAM file containing aligned RNA-sequence reads mapped to *B. malayi*. The second BAM file that contained only unaligned reads were then remapped against the *wBm* genome using Subread-Aligner, resulting in a final BAM file that contained aligned and unaligned RNA-sequence data to *wBm*. Aligned RNA-sequence data in this file was unique to *wBm* only. The aligned RNA-sequence data was visualised alongside *wBm*'s genome in the Artemis genome viewer[190,191], and quantified as FPKM using the Cufflinks program (v. 2.2.1)[130].

The procedure described above was also repeated in reverse, with alignment to the *w*Bm genome conducted first. FPKMs for this 'reverse' alignment were quantified as FPKM using the Cufflinks[130] program, with quantifications compared against the initial alignment for instances of lateral gene transfer. Genes that have a higher FPKM in the reverse alignment as compared to the initial alignment were deemed to have undergone instances of lateral gene transfer.

2.4.5 Annotation of *w*Bm's Genome using RNA-Sequence data

Unannotated regions of *wBm*'s genome were visually inspected for the presence of RNA-sequences aligned to them in the Artemis program[190,191], with their coordinates recorded wherever present. These coordinates were then used to calculate an FPKM value for the regions of interest, and a density plot was generated to visualise the distribution of FPKMs per gene. As a control for false positives, a fifth-percentile threshold was applied to the distribution. If identified regions of interest had an FPKM value above the threshold, an intact ORF greater than 300 base pairs, and/or if a near-full length homologue was found in databases, the region was annotated as a new gene into a GFF file (**A 1-1**). If the region did

not have an intact ORF, but had significant FPKM alignments to the region, then it was recorded as a new pseudogene in a separate GFF file (**A 1-2**).

For analysing and annotating identified regions of interest, Blast2Go[182] was used for the rapid annotation and identification of all genes . This used default parameters, and a pre-formatted sequence database from NCBI (downloaded on 31st May 2016). RFam [184] was also used in order to identify possibly functional RNA domains. Additional genes within the set that had no known function were then analysed by Pfam and InterProScan[122] to identify potential protein domains. Phobus and TMHMM[186,199] were used to more accurately determine the orientation and localisation of the protein, with PSIBlast[200,201] searches used to identify the presence of conserved residues and structures within the proteins of interest. For results found in PSIBlast, alignments were made via the T_Coffee program[202] to formally identify the presence of conserved regions and active sites in the *wBm* protein of interest, which might point to otherwise hidden function. Proteins with noticeable conserved regions had their or a homologue's 3D structure determined via the RCSB Protein data bank (www.rcsb.org[183]). Using the structure, comparisons were made to identify if a fully-conserved active site was present in the *wBm* genes identified, which may provide indications as to the function or role in *wBm*.

The new annotations were then compared to a set of existing annotations generated via automated pipelines to add validity to the results, as well as to illustrate the utility of RNA-sequencing. Specifically, comparisons were made against the automated annotation output by the National Centre for Biotechnology Information's automated annotation pipeline for prokaryotic genomes[120]. The process involved manual examination of gene coordinates between the different annotation files using the Artemis Genome viewer[190,191], identifying regions where the annotation coordinates agreed, overlapped, or disagreed.

2.4.6 Comparison of *w*Bm genes across the *Wolbachia* supergroups

Wolbachia genomes for analysis were obtained from the NCBI's genome database using the search keyword 'Wolbachia'. 2 additional genomes were also obtained from the external site http://nematodes.org/genomes, which maintains the genomes of two parasitic filarial nematodes and their *Wolbachia* endosymbionts.

The Large Scale Blast Score Ratio (LS-BSR)[216] program was used for analysis of the 24 newly identified genes, and the *w*Bm genome as a whole.

2.4.7 Variant calling in the *w*Bm genome

Any generated BAM files of interest to be used for variant calling were first checked to ensure all mate-related flags had been correctly filled in so as to avoid errors. This was done by using the program SAMTools (version 1.9), to first name-sort all the aligned reads using the *sort* command with option *-n*, followed by the *fixmate* command with option *-m*. Following this, duplicates reads within the alignment files were then removed, first by position-sorting all reads using the command *samtools sort*, followed by the command *samtools markdup -r*.

Using these sorted, duplicate-removed BAM files, genotype likelihoods were then predicted using as input all of the generated alignment files against the *w*Bm genome. This used the program BCFTools (version 1.9), with the command *mpileup*, with options *-C 50 -q 10 -Q 15 -E -f*.

With these genotype likelihoods, variants were then called using the BCFTools' *call* command, with options *-c -p 0.005* to generate a initial variant call format file. This was then filtered for quality using BCFTools' vcfutils perl script, with options *-Q 40 -d 30 -w 3 -W 10 -a 1 -p*.

The quality filtered, called variants were then visualised against the *w*Bm genome using the Integrated Genome Viewer program (version 2.3).

Chapter 3 Differential Expression of L3-L4 *B. malayi* and *w*Bm

3.1 Introduction

The rise of omics technology in the past few years has permitted new insights into the biology of pathogenic organisms, particularly those that harbour endosymbionts, leading to a further understanding of basic biology or identification of targets that can be exploited for treatment as, reviewed by Murfin *et al.*[246]. With respect to parasitic filarial nematodes, much of the biochemical interactions between the *Wolbachia* endosymbiont and nematode host have been revealed relatively recently due to the development of such omics technology. For instance, it has been predicted based on comparative genomics studies between the genomes of *w*Bm and *B. malayi*, that the endosymbiont may supplement essential metabolites to the nematode host during periods of high metabolic demands[111,112]. One such metabolite group is nucleotides, possibly inclusive of ATP, as the *w*Bm endosymbiont maintains the full *de-novo* biosynthesis pathway, whilst *B. malayi* maintains only genes for nucleotide interconversion and salvage[111,112]. Further examples include haem[111,247], riboflavin, and FAD, all critical co-factors for multiple enzymatic reactions that the nematode host is incapable of synthesising[112]. This supply-and-demand relationship could explain several of the deleterious effects seen in nematodes upon removal of the *Wolbachia* endosymbiont. Examples of these effects include a cessation of embryogenesis in reproductive adults, an inability for nematodes to successfully develop through the larval stages and a significantly reduced lifespan of the nematode host[80,81,248].

These predictions, and experimentally observed effects of *Wolbachia* depletion on the biology of the nematode host suggests a complex symbiotic relationship. Multiple studies in recent years have attempted to confirm or elucidate the specific mechanisms that underpin this relationship via transcriptomic or proteomic analysis. Such studies have focused mostly on adult life-cycle stages of the nematode and/or its *Wolbachia* endosymbiont, utilising microarray or next-generation-based RNA-sequencing and proteomics analysis of nematodes and *Wolbachia*[166,168,169,249,250].

Specific examples include work done by Darby *et al.*[142] in the parasitic veterinary nematode *Onchocera ochengi*, a closely related bovine parasite of the human parasitic nematode *Onchocerca volvulus*. This study simultaneously sequenced the genome of the associated *Wolbachia* endosymbiont (*wOo*), whilst conducting transcriptomic and proteomic analysis between whole-male nematodes, and separate female body and reproductive tissues. This revealed the smallest *Wolbachia* genome to date at 0.96 Mbp, with 88 genes lost in the *wOo*

strain as compared to previously sequenced strains[142], with the loss of these genes effectively pseudogenising biosynthetic pathways for FAD and riboflavin that were associated with *w*Bm-*B. malayi* symbiosis. The nucleotide biosynthetic pathway is still intact in *wOo* and absent in *O. ochengi*, with additional evidence for the role of ATP provisioning on the part of *wOo* to *O. ochengi*, thus strengthening the significance of this pathway[142].

Another example of such omics analysis has been applied to *Dirofilaria immitis* by Luck *et al.*[171]. *D. immitis* is the causative agent of dog heartworm. Luck *et al.*[171] was able to exploit the relatively large size of *D. immitis* to conduct tissue-specific transcriptomics and proteomics of adult nematodes and their *Wolbachia* endosymbiont (*wDi*), with specific tissues including nematode body walls, digestive tracts, and reproductive tissues, separated by gender. This work echoes previous predictions and findings, with *wDi* overexpressing genes related to transport mechanisms, as well as nucleotide and haem biosynthesis within female uterus tissues as compared to other female or male testis tissues, despite an overall lesser read coverage in the former[171].

Many of the life cycle stages studied so far in relation to *Wolbachia*-nematode symbiosis surrounds L4 to adult life cycle stages, as described earlier and reviewed by Grote *et al.*[251]. By comparison, one important life cycle phase that has received relatively little attention, is the larval development of nematodes and their *Wolbachia* endosymbiont upon infection of the mammalian host. For instance, within a week of *B. malayi* parasitizing its final host, *w*Bm populations are known to rapidly increase to expand throughout the lateral chords of the nematode[72], resulting in an approximately 600-fold population increase[15] (**Figure 3-1**). The key triggers and processes of how *w*Bm can achieve this rapid population expansion are unknown despite this population increase being critical for nematode development. Clearance of *w*Bm at, or before this stage is known to arrest nematode developmental progression to reproductive adults[80,248], with this phenotype being replicated across multiple different filarial nematodes, such as *Brugia malayi*, *B. pahangi*, and *Dirofilaria immitis*[248].

As such, the objective of this chapter's study was to look at the transcriptomic profile of *B. malayi* and its *w*Bm endosymbiont during the first two weeks of infection of the mammalian host (**Figure 3-1**). For this, we generated RNA-sequence data spanning L3 and L4 developmental stages from 3, 7, 11, and 14 days after *B. malayi* infection of the Mongolian gerbil. This 'total' transcriptome from these 4 time-points was taken forward and partitioned into 2 datasets by alignment to their respective genomes[111,112]; **(i)** *B. malayi*-unique reads plus lateral gene transfer events, and **(ii)** *w*Bm-unique reads. Doing so allows for analysis of

the two organism's transcriptomes side-by-side, and subsequent identification of differentially expressed genes and pathways that may contribute to *w*Bm population proliferation and the symbiotic relationship. These 4 time-points were chosen as they span the L3 to L4 developmental moult of the nematode, occurring between days 7 and 11, and as mentioned previously, the *Wolbachia* population is known to undergo a significant population expansion within the first 7 days after infection[15]. Transcriptomic investigation of these time-points will help to further elucidate the relationship between *B. malayi* and *w*Bm.

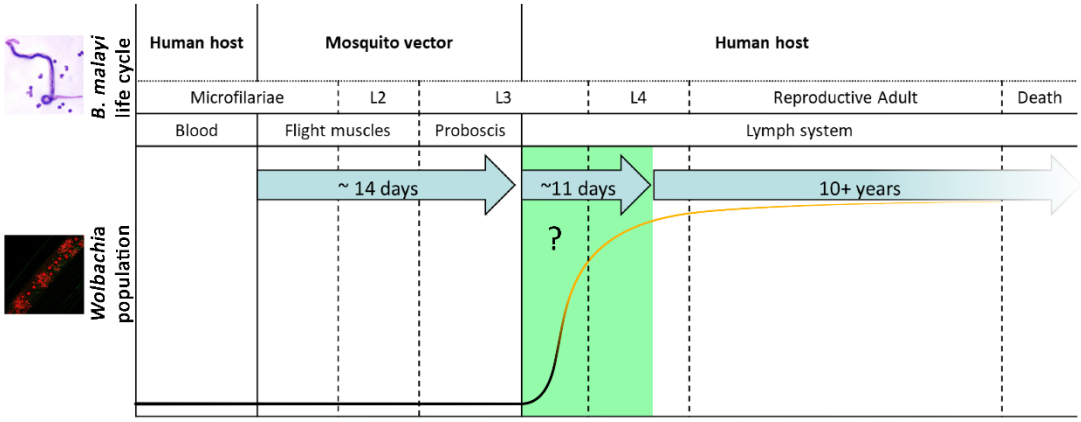


Figure 3-1: Graphical representation of the life cycle of* B. malayi*, and the concurrent population changes with its* Wolbachia *endosymbiont, not drawn to scale. *Developmental life cycle stage of interest for this study is highlighted in green. Note how the* Wolbachia *population increases at an exponential rate from a static population within the first few days of* B. malayi *infection into the final host. Images of* Wolbachia *taken from Taylor et al.*[67]*, and* B. malayi *taken from the Centers for Disease Control and Prevention*[70]

3.2 Results

3.2.1 Overview of Transcriptome alignments

All RNA extraction work in this section was done by Dr Christina Bronowski, with processing and sequencing done by the Liverpool Centre for Genomic Resources

Total RNA sequencing from *B. malayi* nematodes extracted 3, 7, 11 and 14 days post-infection of Mongolian jird hosts yielded a minimum of 16 million trimmed unaligned reads for all replicates, with the distribution of reads displayed in full in **Appendix 2**. With each replicate, a minimum of 89.17% of the total reads mapped to the *B. malayi* host genome- a figure broadly in line with similar transcriptomics experiments published previously[169,171]. The remaining unmapped reads were then extracted and remapped to the *w*Bm genome. As expected, the number of *w*Bm reads increased in line with the expected expansion in *Wolbachia* population rising to an average of 65.70% at 14 days post infection (**Figure 3-2**). Reads that were successfully aligned to either genome were quantified using FeatureCounts[252] for downstream differential expression analysis.

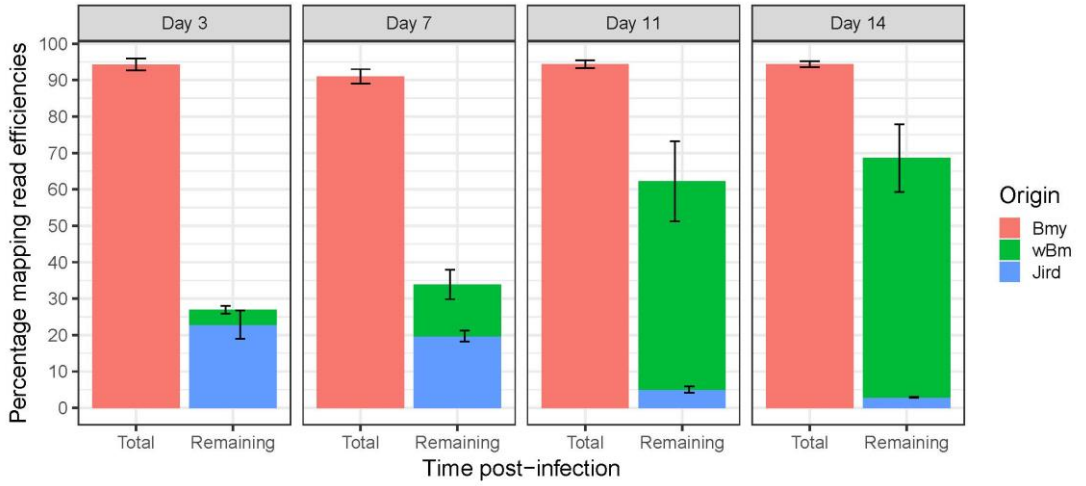


***Figure 3-2**: **Chart to show average mapping percentage efficiencies of reads to the genomes of either** **B. malayi, the w**Bm **endosymbiont, or the gerbil animal host.** Mapping efficiency to the* B. malayi *genome is based on total reads mapping to the* B. malayi *genome, whereas mapping efficiency to the w*Bm *genome is based on the remaining unmapped reads that successfully mapped to the w*Bm *genome. Note the low mapping efficiency of w*Bm *reads during days 3 and 7, as compared to the increased mapping efficiency between days 11 and 14.*

3.2.2 q-PCR confirmation of *w*Bm load

All qPCR work in this section was done by Dr Christina Bronowski, with processing and sequencing done by the Liverpool Centre for Genomic Resources

To confirm that *w*Bm populations increase during the two weeks of study, qPCR was conducted on 10 biological replicates of single nematodes from each of the 4 studied time-points. Briefly, DNA was extracted from 10 individual L3 or L4 larvae from each of the 4 studied time-points, and used for qPCR amplification with primers designed to target the *Wolbachia* surface protein (*wsp*) and *B. malayi* Glutathione-S Transferase (*gst*) genes, as described in McGarry et. al.[15]. Amplifications for each biological replicate were performed in quadruplicate via a Taqman duplex assay, using the FAM fluorophore for the *wsp* readout, and the VIC fluorophore for the *gst* readout, alongside a standard curve for both genes generated by using 7 serial 1-in-10 dilutions of a known stock concentration.

As expected, and mirroring results observed from read mapping efficiencies, qPCR results showed that *w*Bm populations increase as time progresses after infection (**Figure 3-3**).

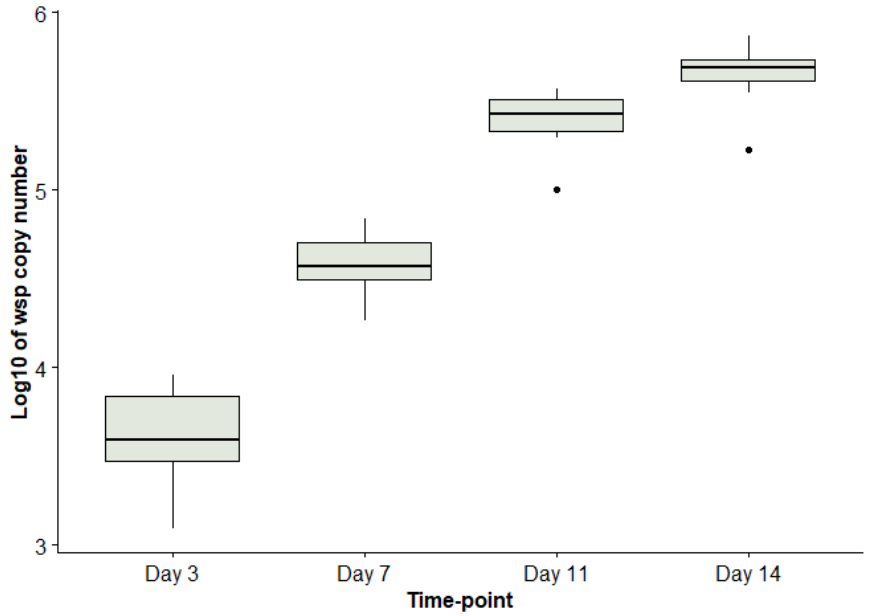


Figure 3-3: Boxplot of* wsp *gene copy number from 10 individual* B. malayi *nematodes collected from each time-point 3, 7, 11, and 14 days post-infection.

3.2.3 Overview of Differential Expression analysis

Pairwise differential expression analysis was conducted with the four time-points of *B. malayi* development. This was done using the EdgeR package[127] in two modes- firstly a time-course, or baseline comparison format using Day 3 as the baseline (Day 3 versus Day 7, Day 3 versus Day 11, Day 3 versus Day 14). This method was used to identify any overall transcriptomic trends that may exist in either *B. malayi* or *w*Bm. The second mode utilised an 'all-versus-all' comparison, with focus on using a time-course comparison format (Day 3 versus Day 7, Day 7 versus Day 11, Day 11 versus Day 14). This was done to take advantage of the increasing read-mapping efficiencies of *w*Bm over time, as well as to corroborate specific changes of interest identified from the baseline comparisons. The full results of both analysis methods are included in **Appendix 2**. Transcriptomics studies focusing on similar nematodes typically focus on an all-versus-all approach in their analysis[170,171,253], likely due to these studies comparing significantly different life cycle stages to one another in isolation. As this study is focusing on changes that occur over a short time span and across closely related life cycle stages, we opted to focus on results from the baseline comparison format to observe any biological changes, as well as the clarity this format brings in terms of changes over time. The EdgeR program was chosen for this purpose due to its widespread use, and available external wrappers for ease of use.

The results of this second analysis corroborate the changes observed during the baseline comparison format, and identifies several additional genes achieving statistical significance. Examples of such newly identified statistically significant genes include Pyrimidine deaminase

and reductase (Day 7 versus Day 14, *wBm0026*), several subunits of a Na+/H+ antiporter system (Day 7 versus Day 14, *wBm0050*, *wBm0389*, *wBm0421*), and the DNA segregation protein FtsK (Day 7 versus Day 11 or Day 14, *wBm0644*). Full results of this analysis is included in **Appendix 2**.

A total of 373 genes of *w*Bm (out of a total of 866) were identified as having statistically significant differential expression across any single pairwise comparison (FDR <0.05). During this study, the magnitude of fold-change ranged from -7 to 11 on a $\log_2$ scale, whilst the number of statistically significant genes ranged from 10 (Day 3 versus Day 7) to 271 (Day 7 versus Day 14).

Some of the genes that showed the highest level of upregulation include the outer surface protein wsp (*wBm0284)*, the Type IV secretion subunit VirD4 (*wBm0283*), and several hypothetical proteins (*wBm0006*, *w*Bm*0043*). This may indicate *w*Bm's prioritisation of structural and transport components during this time, possibly in relation to its growth. Interestingly, the number of genes that were statistically significantly up- or down-regulated during any of the analysed pairwise comparisons was approximately equal (**Table 3-1**). This suggests that there is a more complex mechanism linked to *w*Bm's transcriptome, rather than a simple unidirectional increase in gene expression as the population expands.

wBm	Time-point Comparison	Upregulated genes	Downregulated genes	Total
	Day 3 vs Day 7	4	6	10
	Day 3 vs Day 11	50	24	74
	Day 3 vs Day 14	72	40	112
	Day 7 vs Day 11	90	73	163
	Day 11 vs Day 14	66	76	142
	Day 7 vs Day 14	152	119	271
	Unique genes that show differential expression			373
B. malayi	Time-point Comparison	Upregulated genes	Downregulated genes	Total
	Day 3 vs Day 7	2526	2513	5039
	Day 3 vs Day 11	2407	2468	4875
	Day 3 vs Day 14	2426	2582	5008
	Day 7 vs Day 11	2352	2409	4761
	Day 11 vs Day 14	1164	1292	2456
	Day 7 vs Day 14	1933	1989	3922
	Unique genes that show differential expression			7922

Table 3-1: Distribution of differentially expressed genes identified by the program EdgeR within w*Bm *and* B. malayi *across the studied four time-points. *Note how the number of up- and down-regulated genes are broadly similar to each other in the w*Bm *dataset, indicating no unidirectional increase that can be ascribed solely to population growth.*

By contrast 7,922 unique genes were identified as having statistically significant differential expression in *B. malayi* (out of a total 11,535) across any single pairwise comparison (FDR < 0.05). The magnitude of fold-changes in this dataset ranged from -11 to 13 on a $\log_2$ scale, whilst statistically significant genes ranged from 2,456 (Day 11 versus Day 14), to 5,039 genes (Day 3 versus Day 7, **Table 3-1**).

As *B. malayi* is expected to undergo significant morphological changes from the L3 to the L4 stage, including sexual organ differentiation and moulting, the initial large number of statistically significant genes is to be expected (**Table 3-1**). For example, during the Day 3 versus Day 7 comparison a total of 44 genes annotated as related to the nematode cuticle or cuticular collagen were noted to be upregulated, with some showing a $\log_2$ fold-change of up to 13 (**Appendix 2**). Some of the most highly upregulated examples include *Bm1_11115*, *Bm1_13325*, *Bm1_25060*, and *Bm1_53285*. These same genes were noted to have reduced levels of upregulation during the Day 3 versus Day 11 comparisons (maximum $\log_2$ fold-change of 2), with some considered downregulated during the Day 3 versus Day 14 comparison (down to $\log_2$ fold-change of -0.5, **Appendix 2**). This regulation pattern likely corresponds to the nematode's preparations and subsequent completion of the L3 to L4 moult, with similar observations being noted in previous studies[249,250].

3.2.4 Analysis of consensus between differential expression analysis techniques: EdgeR, DESeq2, CuffDiff

While this work is not intended to act as a benchmarking comparison study for differential expression analysis tools, due to the discovery nature of the work we adopted an additional consensus approach to look for genes that show statistical significance across multiple different algorithms. As such, additional differential expression analysis was carried out with the programs CuffDiff (in combination with CuffLinks, and referred to as CD[129,130,189]) and DESeq2 (referred to as DS[128]), and compared against existing results from EdgeR (referred to as ER[127]). These three programs are some of the most commonly cited differential expression analysis programs available. All three algorithms take an initial assumption that no genes are differentially expressed, but take different approaches in normalisation and modelling to assess if a gene shows statistically significant differential expression between two given conditions.

In brief, CD takes in a given alignment file and corresponding annotation, before counting the number of transcript 'fragments' that map to each gene. From this, the variability in fragment count per gene and per isoform is then modelled across all given replicates using a negative binomial model[129,130]. In the case of isoforms where present, a measure of

uncertainty is also estimated due to the presence of alternative splicing generating ambiguously mapped reads. A combination of these estimates, and replicate variability are then combined into a beta negative binomial model to estimate count variance for each transcript within the given genome. These variance estimates per transcript are then used as the basis to determine statistically significant differential expression per transcript.

On the other hand, DS takes in a matrix of read counts from an alignment file and associated genome calculated by an external program such as FeatureCounts[252]. From this, DS calculates gene-wise maximum likelihood estimates and $\log_2$ fold-changes, before a fitted curve is generated to estimate the amount of variance present amongst given replicates. An empirical Bayes procedure is then used to 'shrink' the amount of variance present for each gene to control for genes with low read counts, reducing bias in fold-changes. A second curve is then fitted to this 'shrunken' data, before conducting a Wald's significance test, with the null hypothesis that genes are not differentially expressed[128].

Finally, like DS, ER takes in a matrix of read counts calculated by an external program[127]. From this, the variance level of each gene across all condition replicates is estimated by an empirical Bayes method, and a negative binomial model is generated for the spread of read counts per gene per sample, which can be reduced to a Poisson model if the variance level per gene is 0. From these values, differential expression of particular genes is then calculated by an exact test analogous to Fisher's exact test, with the null hypothesis that genes are not differentially expressed[127].

Differential expression analysis was conducted with the two new algorithms CD and DS using an all-versus-all comparison on both *w*Bm and *B. malayi*. Following this, the results of all three algorithms were compared to identify statistically significant genes which are shared across multiple algorithms (**Figure 3-4**, **Figure 3-5**). Previous benchmarking studies[254] have indicated that CD is less sensitive, but also less prone to Type I errors whereas DS and ER are more sensitive, but also more prone to Type I errors, and thus suitable for more exploratory analysis[254]. Thus, genes that are identified by CuffDiff and at least one other program could be considered as a reliable result.

Surprisingly, when Day 3 was utilised as a baseline we observed CD identifying significantly more genes on its own as compared to ER or DS (**Figure 3-4**), especially during early time-point comparisons (Day 3 versus Day 7). This is possibly due to the comparatively low read counts obtained from samples on Day 3 as compared to other studied time-points affecting CD's statistical tests. Despite this, we see a total of 70 and 85 genes that were identified by

CuffDiff and at least one other analysis program when comparing Day 3 to Day 11 or Day 14 respectively (**Figure 3-4, Table 3-2**). This trend is reversed when Day 7 or Day 11 is utilised as a baseline, where we observed expected behaviour amongst the three algorithms (**Figure 3-4**). Namely, that a comparatively large number of genes were identified as significantly expressed by ER and DS, and a lesser number of genes by CD (**Figure 3-4**). This is exemplified by the Day 7 versus Day 14 comparison that shows all significantly expressed genes identified by DS as also being identified by ER. This comparison also has the highest number of statistically significant differentially expressed genes out of all comparisons done during this study (65 genes by consensus, **Figure 3-4**, and 271 genes by EdgeR alone, **Table 3-1**). This may be reflective of the different 'states' that *w*Bm may be observed in, i.e. the highly replicative state during the initial L3 infection, and the slower replicative state seen in L4 nematodes and onwards.

In the nematode host, the interactions between all three algorithms is as expected from previous benchmarking studies, with CD identifying the fewest genes amongst all three algorithms regardless of the comparison conducted (**Figure 3-5**). This is followed by ER, with DS yielding the highest number of differentially expressed genes, with these two algorithms having the highest overlap in terms of identified genes amongst the three. Interestingly, when using Day 3 as a baseline the three algorithms frequently identified high numbers of genes as being differentially expressed, at over 3,100 genes per comparison. This is likely reflective of the high read counts available for the *B. malayi* transcriptome, as well as the significant developmental changes the nematode undergoes through the two weeks studied. As an example, of the 44 genes related to nematode cuticle or cuticular collagen identified as upregulated by ER alone, 21 of these genes were also identified by all three algorithms. These identified genes have a maximum $\log_2$ fold-change of 6.7, with examples being *Bm1_00815*, *Bm1_37110*, *Bm1_11865*, and *Bm1_10410*.

This set of overlaps will help in further analysis of the *w*Bm-*B. malayi* transcriptome by bringing forward the most 'reliable' genes in terms of differential gene expression.

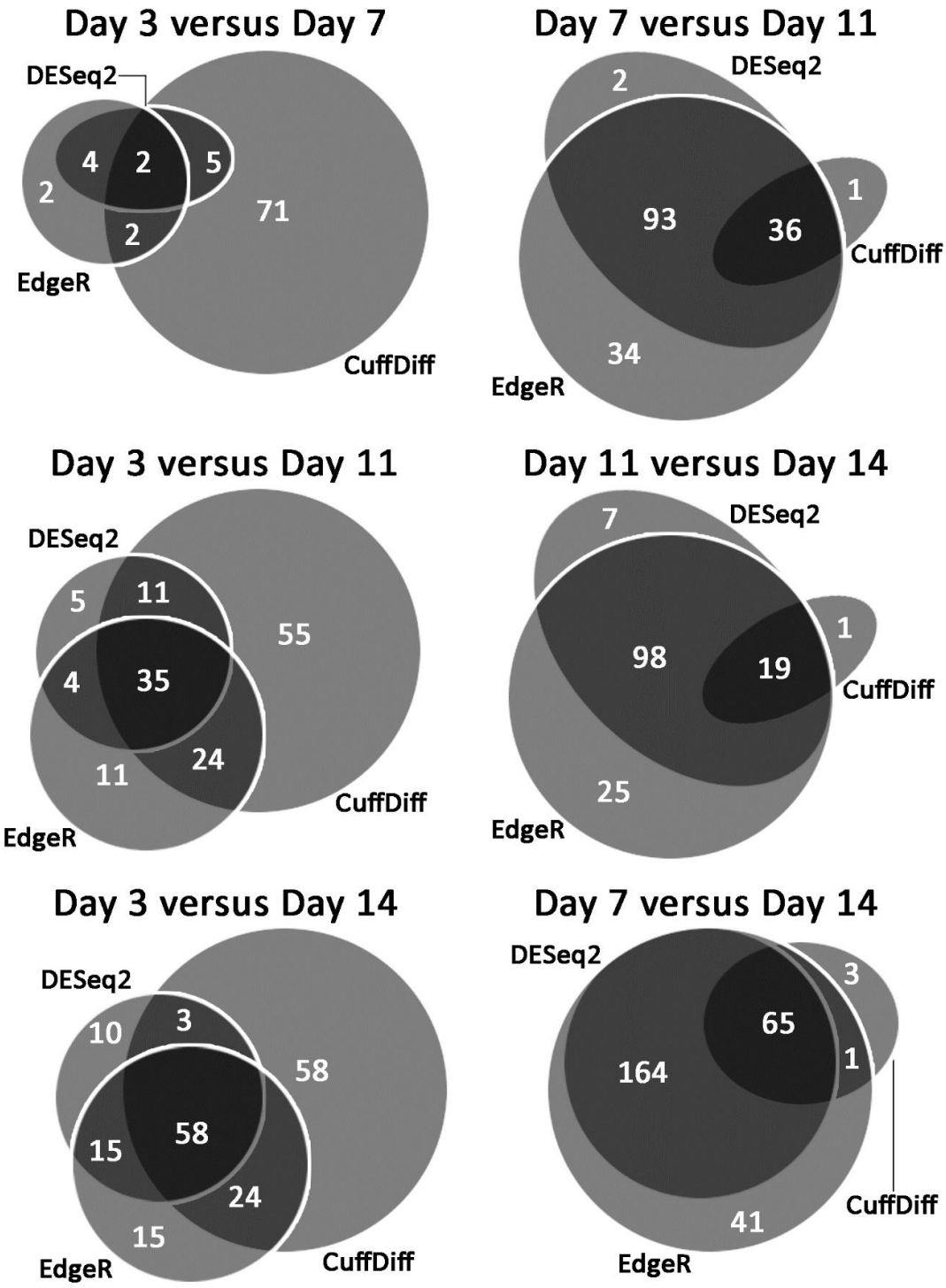


***Figure 3-4**: **Comparison of statistically significant genes identified from three differential expression analysis algorithms on the w**Bm **dataset.** Note the number of genes that EdgeR and DESeq2 frequently share. Also note how CuffDiff frequently has high numbers of 'unique' significant genes when Day 3 is included as the baseline.*

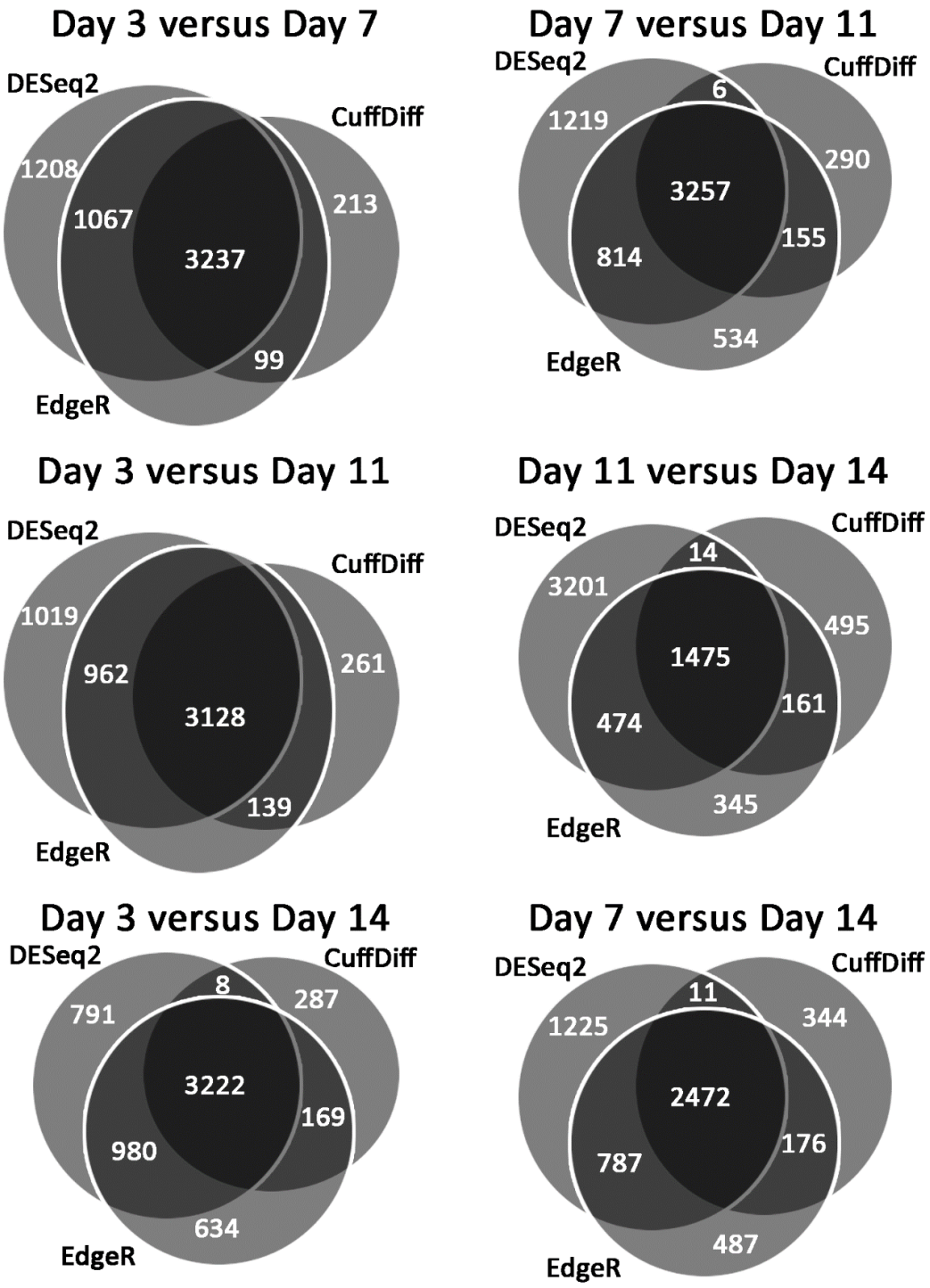


Figure 3-5: Comparison of statistically significant genes identified from three differential expression analysis algorithms on the* B. malayi *dataset. *Note again the large number of shared genes between EdgeR and DESeq2, and also how DESeq2 has high numbers of 'unique' significant genes.*

D3vD7 (9 genes)	D3vD11 (70 genes)		D3vD14 (85 genes)	
wBm0138	*wBm0005*	*wBm0521*	*wBm0001*	*wBm0443*
wBm0314	*wBm0006*	*wBm0537*	*wBm0005*	*wBm0454*
wBm0350	*wBm0007*	*wBm0541*	*wBm0006*	*wBm0475*
wBm0387	*wBm0008*	*wBm0544*	*wBm0007*	*wBm0484*
wBm0403	*wBm0019*	*wBm0567*	*wBm0008*	*wBm0490*
wBm0495	*wBm0028*	*wBm0583*	*wBm0013*	*wBm0494*
wBm0508	*wBm0051*	*wBm0596*	*wBm0019*	*wBm0496*
wBm0772	*wBm0053*	*wBm0613*	*wBm0028*	*wBm0500*
wBm0786	*wBm0089*	*wBm0616*	*wBm0043*	*wBm0508*
	wBm0092	*wBm0630*	*wBm0051*	*wBm0521*
	wBm0097	*wBm0637*	*wBm0053*	*wBm0537*
	wBm0124	*wBm0670*	*wBm0087*	*wBm0538*
	wBm0138	*wBm0674*	*wBm0089*	*wBm0544*
	wBm0146	*wBm0704*	*wBm0092*	*wBm0567*
	wBm0179	*wBm0708*	*wBm0124*	*wBm0572*
	wBm0181	*wBm0722*	*wBm0146*	*wBm0583*
	wBm0206	*wBm0727*	*wBm0154*	*wBm0596*
	wBm0210	*wBm0755*	*wBm0179*	*wBm0606*
	wBm0211	*wBm0767*	*wBm0181*	*wBm0609*
	wBm0221	*wBm0771*	*wBm0182*	*wBm0613*
	wBm0227	*wBm0772*	*wBm0193*	*wBm0628*
	wBm0239	*wBm0775*	*wBm0202*	*wBm0630*
	wBm0276	*wBm0786*	*wBm0210*	*wBm0637*
	wBm0284	*wBm0804*	*wBm0211*	*wBm0647*
	wBm0291	*wBm0806*	*wBm0221*	*wBm0659*
	wBm0296	*wBm0808*	*wBm0227*	*wBm0674*
	wBm0320	*wBmNew0001*	*wBm0239*	*wBm0695*
	wBm0355		*wBm0276*	*wBm0704*
	wBm0384		*wBm0284*	*wBm0708*
	wBm0387		*wBm0291*	*wBm0717*
	wBm0399		*wBm0320*	*wBm0738*
	wBm0402		*wBm0335*	*wBm0739*
	wBm0411		*wBm0350*	*wBm0740*
	wBm0424		*wBm0355*	*wBm0767*
	wBm0434		*wBm0362*	*wBm0771*
	wBm0437		*wBm0384*	*wBm0772*
	wBm0443		*wBm0387*	*wBm0774*
	wBm0451		*wBm0399*	*wBm0786*
	wBm0454		*wBm0411*	*wBm0804*
	wBm0484		*wBm0414*	*wBm0806*
	wBm0494		*wBm0426*	*wBm0808*
	wBm0495		*wBm0434*	*wBmNew0001*
	wBm0508		*wBm0437*	

Table 3-2: List of genes identified as differentially expressed across the CuffDiff algorithm and at least one other algorithm (DESeq2 or EdgeR).

3.2.5 Pathway map generation and analysis strategy of the *Wolbachia* endosymbiont

In an attempt to derive additional biological meaning from the results of differential expression, Gene Ontology (GO) enrichment (via the program Blast2GO[182]), or Pathway enrichment (using the Generally Applicable Gene set Enrichment (GAGE) program[135]) were performed on the *w*Bm transcriptome.

To summarise these two techniques, GO terms are a curated 'dictionary' of terms that are assigned to genes, offering a standardised method of describing gene function or localisation. GO enrichment programs, such as one bundled in Blast2GO[182] are designed to take a user-defined set of genes, and compare this versus a background set of genes, before determining if there are GO terms that are noticeably enriched (more abundant as compared to chance), within the user-defined set as can be expected by random chance when picked from the background gene set. By contrast, GAGE[135] overlays all differential expression fold-change data onto a 'background' set of genes, in this case the background is determined by biological pathways as defined by the Kyoto Encyclopaedia for Genes and Genomes (KEGG) pathway database[255]. This overlay is then used to determine if a given gene set shows statistically significant differential expression as compared to background, or possibly if the pathway as a whole is statistically significantly enriched with either up- or down-regulated genes. In this instance, gene sets were defined based on shared significance between the three differential expression programs described earlier (FDR < 0.05).

These automated techniques depend on a suitably large background gene set to conduct statistical analysis, of which *w*Bm does not have. This is due to *w*Bm's evolution as an obligate endosymbiont, which has resulted in generally small genome sizes and many biochemical pathways becoming lost or significantly reduced as compared to model organisms[111,136]. Because of this, these techniques were unlikely to identify many GO terms or pathways that are enriched from background, and indeed upon applying the two techniques to this dataset, no statistically significant GO terms or pathways were identified.

As a result of this, we opted for a non-automated review of the data, via manual curation of biological pathways between both *w*Bm and *B. malayi*. To do so, we incorporated information from the Kyoto Encyclopaedia of Genes and Genomes (KEGG) and the BioCyc Database collections[255–257]. These are publicly available databases that incorporate genome and pathway information on a wide range of organisms, and is based on a mix of public literature, experimental evidence, and pathways known to exist within model organisms[255–257]. As previously observed by their respective sequencing projects[111,112], gaps in a range of

biological pathways were observed in both *w*Bm and *B. malayi*. Examples being glycolysis and amino acid metabolism in *w*Bm, and nucleotide biosynthesis and haem in *B. malayi*. Utilising this pathway knowledge of both organisms, we then took the opportunity to manually reconstruct certain pathways of interest that are predicted to play a role in symbiosis (**Table 3-3**), taking into account the presence of these gaps, and in some cases how the pathways of the two organisms overlap with each other directly. These reconstructed pathways are included in **Appendix 3**.

Pathway name (*KEGG Identifier*)	Total genes	Day 3 versus Day 7	Day 3 versus Day 11	Day 3 versus Day 14
Purine biosynthesis (*wBm00230*)	20	11	14	13
Pyrimidine biosynthesis (*wBm00240*)	15	12	11	10
TCA cycle (*wBm00020*)	18	9	12	9
Glycolysis (*wBm00010*)	12	4	4	4
DNA replication (*wBm03030*)	13	9	7	8
Glutathione Metabolism (*wBm00480*)	5	4	4	4
Haem biosynthesis (*wBm00860*)	12	10	11	11
Type IV secretion system (*wBm03070*)	10	8	8	8

Table 3-3: Several pathways of interest identified in this study containing genes that were consistently upregulated or downregulated over the time-points studied. *The KEGG identifier included for each pathway is based on knowledge gained from model organisms, and includes significant gaps due to w*Bm*'s degenerate genome.*

Following on from this, we then overlaid results from differential expression onto these pathways with the aid of the Pathview program[258], which is designed to map user-defined fold-change differences onto genes in KEGG pathways for analysis. Genes in *w*Bm were considered as upregulated or downregulated if they had a $\log_2$ Fold Change > 0.2, or < -0.2 respectively, according to the default settings of the Pathview analysis program. FDR significance values for each gene was ignored for this mapping, as the objective was to gain a general overview of whether certain pathways were up- or downregulated. As the fold-change of genes were not noted to differ significantly between CD, DS, and ER, we utilised results from ER due to availability of external wrappers for ease of use[259].

3.2.6 Glycolysis and TCA cycle pathways in *Wolbachia* are suggestive of alternating control during early nematode development

Genome sequencing suggests *w*Bm maintains a truncated glycolysis pathway, with the initial steps of extracellular glucose uptake, phosphorylation, and processing into Fructose-1,6-bisphosphate missing from its genome[111]. The role of these enzymes has possibly been supplanted by the host, with localisation of *B. malayi* glycolysis enzymes (specifically fructose-1-6-bisphosphate aldolase, triosephosphate isomerase, L-lactate dehydrogenase,

enolase, glyceraldehyde-3-phosphate dehydrogenase, and phosphoglycerate kinase) to *w*Bm surface proteins having been observed in a previous study[152].

Manual investigation of *w*Bm pathways overlaid with differential expression fold-changes indicated an alternating pattern of regulation in the Glycolysis and TCA cycle pathways during L3-L4 nematode larval development. Out of the 9 genes that *w*Bm maintains, which are responsible for the glycolysis/gluconeogenesis pathway, 5 genes were downregulated across all 3 pairwise comparisons, with 2 additional genes (Fructose bisphosphate aldolase, *wBm0097* and Triosephosphate isomerase, *wBm0408*) being downregulated in at least 1 pairwise comparison (**Table 3-3**, **Figure 3-6-A**). The exceptions to this trend were Fructose-1,6-bisphosphatase (*wBm0132*) and Pyruvate Phosphate Dikinase (*wBm0209*), shown to be upregulated in at least 2 pairwise comparisons, although these enzymes are known to act in the gluconeogenic direction[23,156]. This indicates coordinated downregulation of the glycolysis pathway and is accompanied by genes within the Tricarboxylic Acid Cycle (TCA) being upregulated during later comparisons. These include upregulation of transcripts such as Malate dehydrogenase (*wBm0244*), Citrate synthase (*wBm0735*) and the two Dihydrolipoamide dehydrogenase subunits (*wBm0559 and wBm0561*, **Table 3-3**, **Figure 3-6-C**). Components of the glutaminolysis pathway was also seen to be consistently upregulated during the study. This pathway controls the conversion of glutamine, through to glutamate, then finally 2-oxoglutarate, where it can enter the TCA cycle as an alternative carbon source to pyruvate. Specifically, the genes that encode the Glutamate synthase beta chain subunit, Carbamoyl phosphate synthase large subunit, and Glutamine amidotransferase (*wBm0051*, *wBm0512*, and *wBm0800* respectively), were all noted to be consistently upregulated, whilst *wBm0232* and wBm0654 was noted to be upregulated across two pairwise comparisons (**Figure 3-6-D**). The genes *wBm0512* and *wBm0654* together form the Carbamoyl-phosphate synthetase enzyme, which performs the first committing step in pyrimidine biosynthesis, another pathway which is seen to be highly upregulated in *w*Bm during this time (**Figure 3-6-D**, **Figure 3-7B**).

Additionally, several of the genes identified across the 3 different pathways were noted as statistically significant across CD and at least one other algorithm, lending strength to these observations. These include *wBm0051*, *wBm0097*, *wBm0443*, and *wBm0674* (**Table 3-2**, **Figure 3-6**).

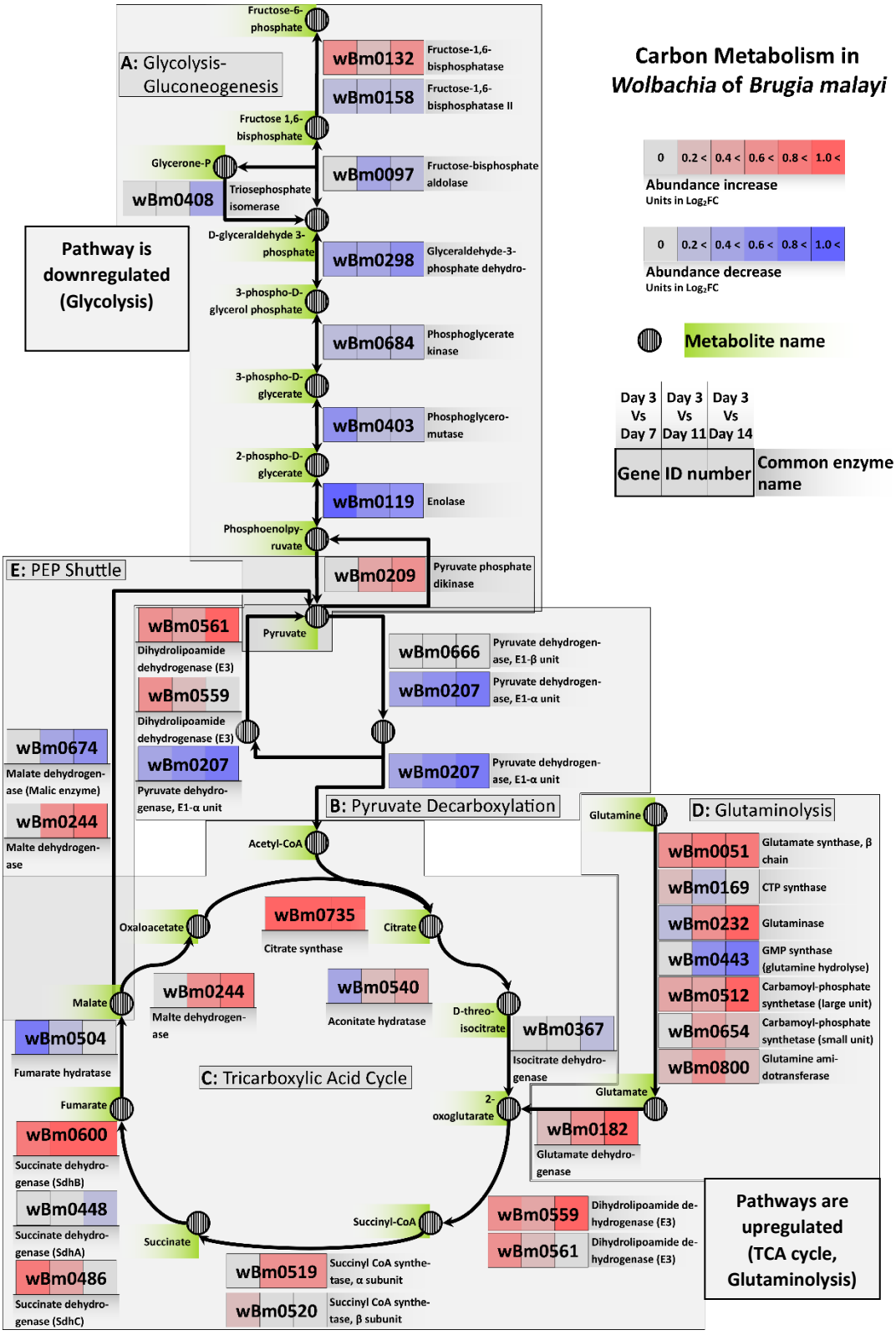


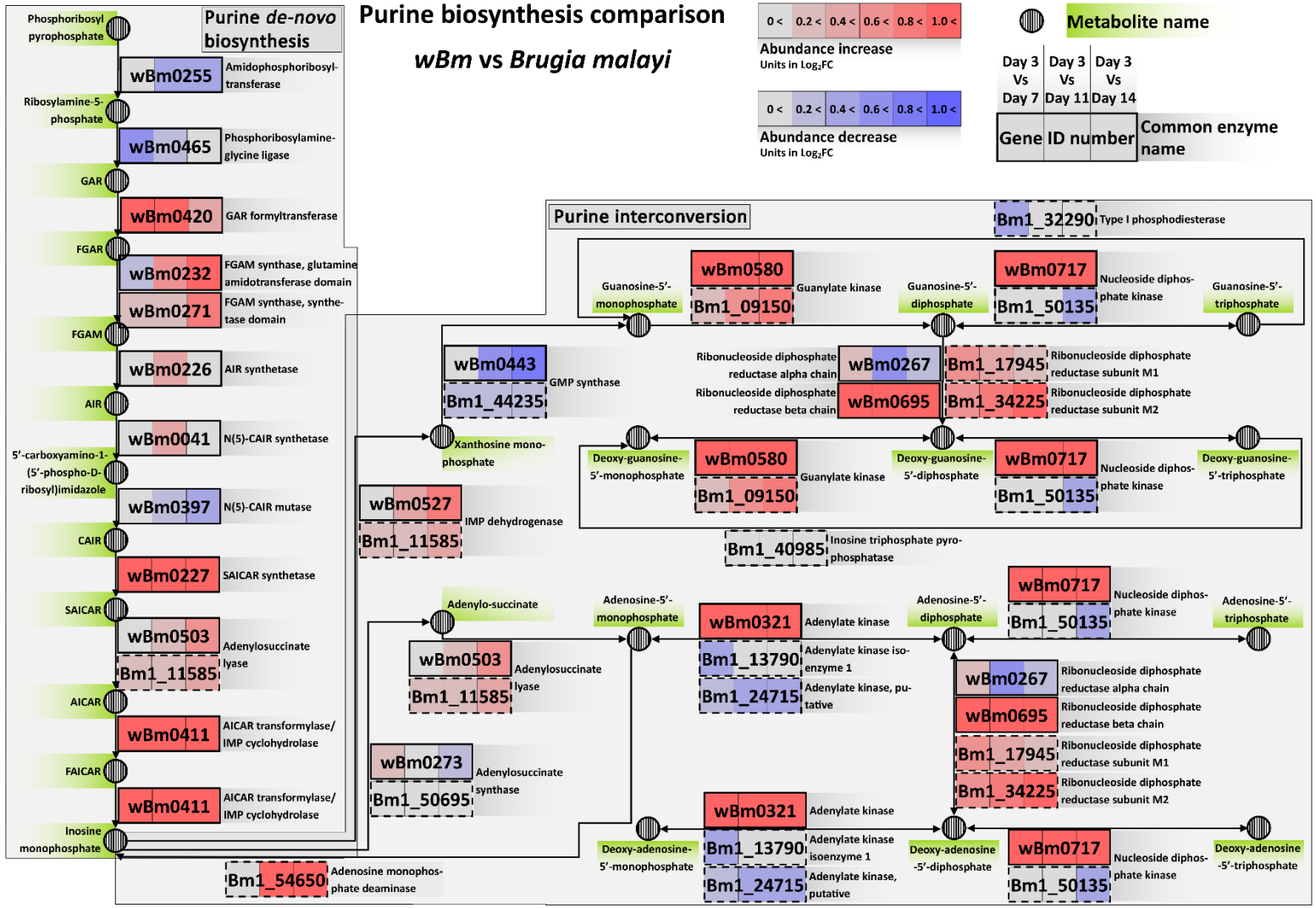
Figure 3-6: Representation of all wBm genes that make up the Glycolysis/Gluconeogenesis, TCA, and Glutaminolysis pathways. *Each large box represents a single enzyme, divided into three smaller boxes to represent the associated time-point comparison. These smaller boxes are colour-coded with varying intensities of red or blue to represent the intensity of up- or down-regulation respectively, see top-right. Note the consistent downregulation of many glycolytic pathway components (**boxes A and B**), and converse upregulation of many components in the TCA and Glutaminolysis pathway (**boxes C and D**). Note also the lack of initial steps for the absorption and breakdown of glucose in the glycolysis pathway.*

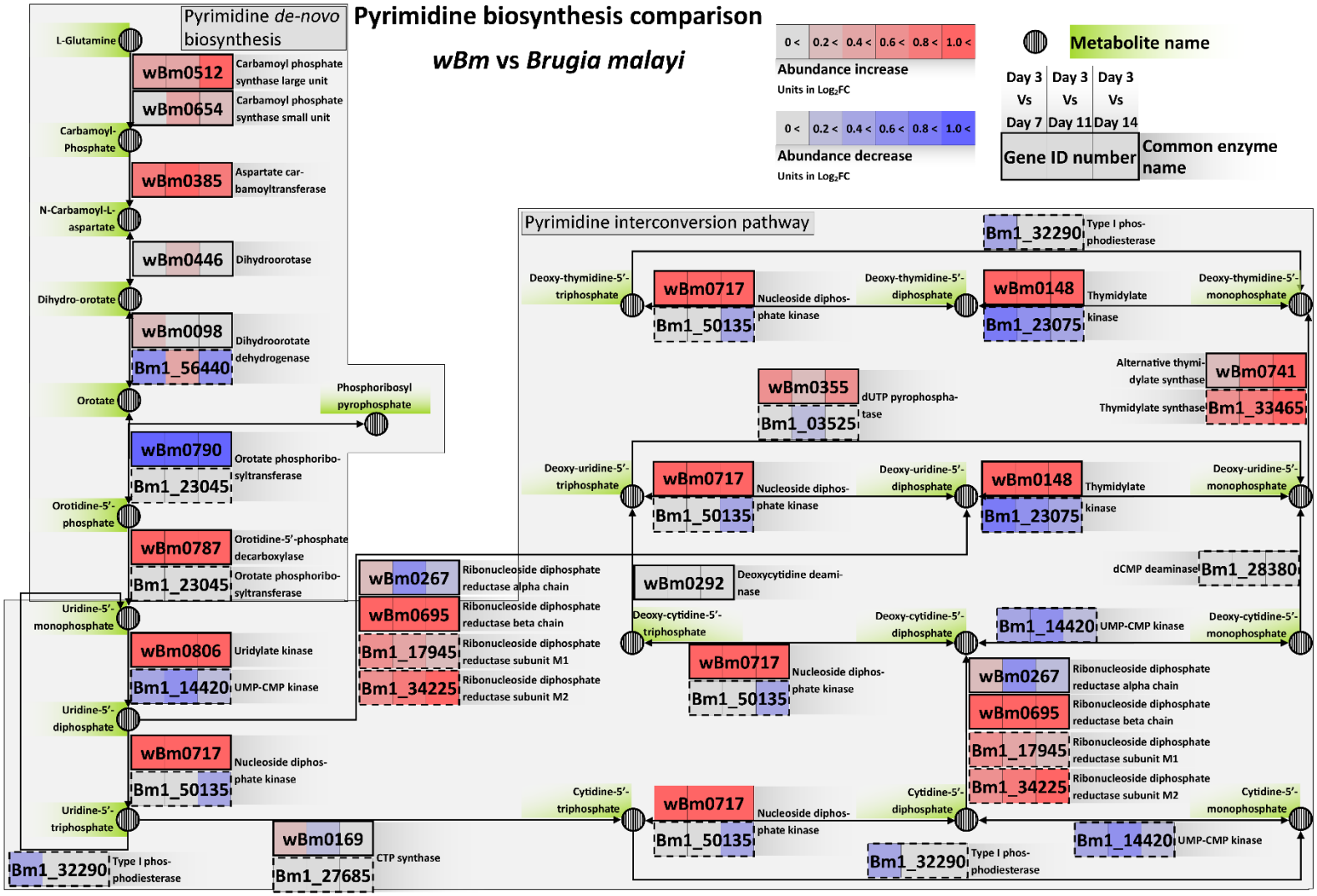
3.2.7 *Wolbachia* nucleotide biosynthesis supports hypothesis of complementing *Brugia malayi* DNA replication

A fully intact pathway for the *de-novo* biosynthesis and interconversion of nucleotides (purines and pyrimidines) is maintained by *w*Bm, in contrast to the nematode host that only maintains pathways for extracellular salvage and interconversion of nucleotides[111]. Differential expression analysis reveals that many components of these pathways were highly upregulated in later time-points, several of which formed committing steps for the pathways when overlaid (*wBm0103*, *wBm0385*, *wBm0420*, *wBm0512*, *wBm0787*) (**Table 3-3, Figure 3-7**). Two additional genes involved in these committing steps showed statistically significant upregulation by consensus of CuffDiff and at least one other algorithm (*wBm0227*, *wBm0411*, *wBm0806* **Table 3-2**). Upregulation of these genes was also accompanied by several of *w*Bm's nucleotide interconversion genes having statistically significant differential expression during several of the analysed time-point comparisons (**Table 3-2**). Examples of statistically significant genes involved in the pathways include *wBm0695*, *wBm0717* (purine metabolism pathway), *wBm0355*, *wBm0695*, *wBm0717* (pyrimidine metabolism pathway).

This consistent upregulation of many components in *w*Bm purine-pyrimidine biosynthesis correlates with the statistically significant upregulation of *Brugia malayi*'s DNA replication pathway during the Day 3 versus Day 14 time-point (**Table 3-4**). Yet when overlaid with *B. malayi*'s degenerated nucleotide biosynthetic pathways (**Figure 3-7A**, gene boxes in dotted lines) there is a noticeable contrast in activity, with *B. malayi* genes showing limited differential expression (examples being *Bm1_44235*, *Bm1_11585*, *Bm1_23045*, *Bm1_14420*), with a few exceptions (examples being *Bm1_34225*, *Bm1_33465*). This correlation between *w*Bm's upregulated *de-novo* nucleotide biosynthesis pathways and *B. malayi*'s upregulated DNA replication pathway continues to occur when analysing later time-points comparison, e.g. Day 7 versus Day 11 or Day 14 (see tables in **Appendix 2**). Several genes also show statistical significance in at least 2 of the 3 algorithms during these later time-point comparisons, with examples including *wBm0226*, *wBm0232*, *wBm411*, *wBm420* and *wBm0527* in the purine biosynthetic pathway, and *wBm0512* in the pyrimidine biosynthetic pathway (**Appendix 2**).

Figure 3-7A and B (next 2 pages): **Combined pathway diagrams illustrating the Purine and Pyrimidine biosynthesis pathways coupled to gene expression data for *B. malayi* and w*Bm* side-by-side.** Brugia malayi *enzymes have a dotted line border, whilst w*Bm *enzymes have a solid border. Note the lack of* Brugia malayi *enzymes in the de-novo synthesis stages of both pathways, yet presence of multiple enzymes for nucleotide interconversions that w*Bm *does not maintain. Also note that nearly all of the w*Bm *enzymes in both the purine and pyrimidine pathways are consistently upregulated, with only a few showing downregulation during later time-points. This contrasts with the few* B. malayi *genes present in these pathways that show little upregulation across the studied pairwise comparisons*


Purine biosynthesis comparison
wBm vs Brugia malayi
Abundance increase
Units in Log2FC
0 < 0.2 < 0.4 < 0.6 < 0.8 < 1.0 <
Abundance decrease
Units in Log2FC
0 < 0.2 < 0.4 < 0.6 < 0.8 < 1.0 <
Metabolite name
Day 3 Vs Day 7
Day 3 Vs Day 11
Day 3 Vs Day 14
Gene ID number
Common enzyme name
Purine de-novo biosynthesis
Phosphoribosyl pyrophosphate
wBm0255
Amidophosphoribosyl-transferase
Ribosylamine-5-phosphate
wBm0465
Phosphoribosylamine-glycine ligase
GAR
wBm0420
GAR formyltransferase
FGAR
wBm0232
FGAM synthase, glutamine amidotransferase domain
wBm0271
FGAM synthase, synthetase domain
FGAM
wBm0226
AIR synthetase
AIR
wBm0041
N(5)-CAIR synthetase
5'-carboxyamino-1-(5'-phospho-D-ribosyl)imidazole
wBm0397
N(5)-CAIR mutase
CAIR
wBm0227
SAICAR synthetase
SAICAR
wBm0503
Bm1_11585
Adenylosuccinate lyase
AICAR
wBm0411
AICAR transformylase/IMP cyclohydrolase
FAICAR
wBm0411
AICAR transformylase/IMP cyclohydrolase
Inosine monophosphate
Bm1_54650
Adenosine monophosphate deaminase
Purine interconversion
Bm1_32290
Type I phosphodiesterase
Guanosine-5'-monophosphate
wBm0580
Bm1_09150
Guanylate kinase
Guanosine-5'-diphosphate
wBm0717
Bm1_50135
Nucleoside diphosphate kinase
Guanosine-5'-triphosphate
wBm0443
Bm1_44235
GMP synthase
Xanthosine monophosphate
Ribonucleoside diphosphate reductase alpha chain
wBm0267
Ribonucleoside diphosphate reductase beta chain
wBm0695
Bm1_17945
Ribonucleoside diphosphate reductase subunit M1
Bm1_34225
Ribonucleoside diphosphate reductase subunit M2
Deoxy-guanosine-5'-monophosphate
wBm0580
Bm1_09150
Guanylate kinase
Deoxy-guanosine-5'-diphosphate
wBm0717
Bm1_50135
Nucleoside diphosphate kinase
Deoxy-guanosine-5'-triphosphate
wBm0527
Bm1_11585
IMP dehydrogenase
Bm1_40985
Inosine triphosphate pyrophosphatase
Adenylo-succinate
Adenosine-5'-monophosphate
wBm0321
Adenylate kinase
Bm1_13790
Adenylate kinase isoenzyme 1
Bm1_24715
Adenylate kinase, putative
Adenosine-5'-diphosphate
wBm0717
Bm1_50135
Nucleoside diphosphate kinase
Adenosine-5'-triphosphate
wBm0503
Bm1_11585
Adenylosuccinate lyase
wBm0267
Ribonucleoside diphosphate reductase alpha chain
wBm0695
Ribonucleoside diphosphate reductase beta chain
Bm1_17945
Ribonucleoside diphosphate reductase subunit M1
Bm1_34225
Ribonucleoside diphosphate reductase subunit M2
wBm0273
Bm1_50695
Adenylosuccinate synthase
Deoxy-adenosine-5'-monophosphate
wBm0321
Bm1_13790
Bm1_24715
Adenylate kinase
Adenylate kinase isoenzyme 1
Adenylate kinase, putative
Deoxy-adenosine-5'-diphosphate
wBm0717
Bm1_50135
Nucleoside diphosphate kinase
Deoxy-adenosine-5'-triphosphate


Pyrimidine biosynthesis comparison
wBm vs Brugia malayi
Pyrimidine de-novo biosynthesis
L-Glutamine
wBm0512
Carbamoyl phosphate synthase large unit
wBm0654
Carbamoyl phosphate synthase small unit
Carbamoyl-Phosphate
wBm0385
Aspartate carbamoyltransferase
N-Carbamoyl-L-aspartate
wBm0446
Dihydroorotase
Dihydro-orotate
wBm0098
Dihydroorotate dehydrogenase
Bm1_56440
Orotate
Phosphoribosyl pyrophosphate
wBm0790
Orotate phosphoribosyltransferase
Bm1_23045
Orotidine-5'-phosphate
wBm0787
Orotidine-5'-phosphate decarboxylase
Bm1_23045
Orotate phosphoribosyltransferase
Uridine-5'-monophosphate
wBm0806
Uridylate kinase
Bm1_14420
UMP-CMP kinase
Uridine-5'-diphosphate
wBm0717
Nucleoside diphosphate kinase
Bm1_50135
Uridine-5'-triphosphate
Bm1_32290
Type I phosphodiesterase
wBm0169
CTP synthase
Bm1_27685
wBm0267
Ribonucleoside diphosphate reductase alpha chain
wBm0695
Ribonucleoside diphosphate reductase beta chain
Bm1_17945
Ribonucleoside diphosphate reductase subunit M1
Bm1_34225
Ribonucleoside diphosphate reductase subunit M2
0 < 0.2 < 0.4 < 0.6 < 0.8 < 1.0 <
Abundance increase
Units in Log2FC
0 < 0.2 < 0.4 < 0.6 < 0.8 < 1.0 <
Abundance decrease
Units in Log2FC
Metabolite name
Day 3 Vs Day 7
Day 3 Vs Day 11
Day 3 Vs Day 14
Gene ID number
Common enzyme name
Pyrimidine interconversion pathway
Deoxy-thymidine-5'-triphosphate
wBm0717
Nucleoside diphosphate kinase
Bm1_50135
Deoxy-thymidine-5'-diphosphate
wBm0148
Thymidylate kinase
Bm1_23075
Deoxy-thymidine-5'-monophosphate
Bm1_32290
Type I phosphodiesterase
wBm0355
dUTP pyrophosphatase
Bm1_03525
Alternative thymidylate synthase
wBm0741
Thymidylate synthase
Bm1_33465
Deoxy-uridine-5'-triphosphate
Deoxy-uridine-5'-diphosphate
Deoxy-uridine-5'-monophosphate
wBm0292
Deoxycytidine deaminase
dCMP deaminase
Bm1_28380
Deoxy-cytidine-5'-triphosphate
Deoxy-cytidine-5'-diphosphate
Bm1_14420
UMP-CMP kinase
Deoxy-cytidine-5'-monophosphate
Cytidine-5'-triphosphate
Cytidine-5'-diphosphate
Cytidine-5'-monophosphate

3.2.8 *Wolbachia* DNA replication machinery indicates complementation of bacterial population dynamics

The DNA replication machinery of *w*Bm is streamlined, as compared to other model bacteria, and consists of the DNA polymerase III holoenzyme, made up of the ε-α catalytic polymerase subunit and γ /τ-δ'-δ complex, as well as various other enzymes and clamps such as DNA helicase, RNAse H, DNA ligase, SSB and β clamps (**Figure 3-8**). The DNA polymerase III complex are responsible for performing DNA replication itself[260,261], with additional mentioned enzymes playing a dual role in DNA replication and DNA repair, and possibly mRNA turnover[262–264].

During early stages of nematode infection into a host, multiple enzymes outside of the ε-α catalytic subunit and γ /τ-δ'-δ complex show upregulation across all studied time-points (**Figure 3-8**). As time progressed, multiple components of the ε-α catalytic subunit and γ /τ-δ'-δ complex showed downregulation, in conjunction with several other enzymes (**Figure 3-8**). This includes the alpha (*wBm0499*), gamma/tau (*wBm0434*), and delta-prime (*wBm0639*) subunits of the complex, of which the gamma/tau complex shows statistically significant differentiation across CuffDiff and at least one other algorithm (**Table 3-2**).

This indicates that between Days 3 and 7, *w*Bm DNA replication rates are highly active, before being downregulated from day 7 onwards. This is reflective of *w*Bm population dynamics, which show a large population increase during the initial week, followed by reduced population growth rate in the subsequent week of infection. Indeed, when analysing the Day 7 vs Day 11 or Day 14 comparisons, multiple genes in the key ε-α catalytic subunit and γ /τ-δ'-δ complexes continue downregulation, with *wBm0434* (encoding the polymerase γ /τ subunit) showing statistically significant downregulation across all 3 algorithms used (see tables in **Appendix 2**).

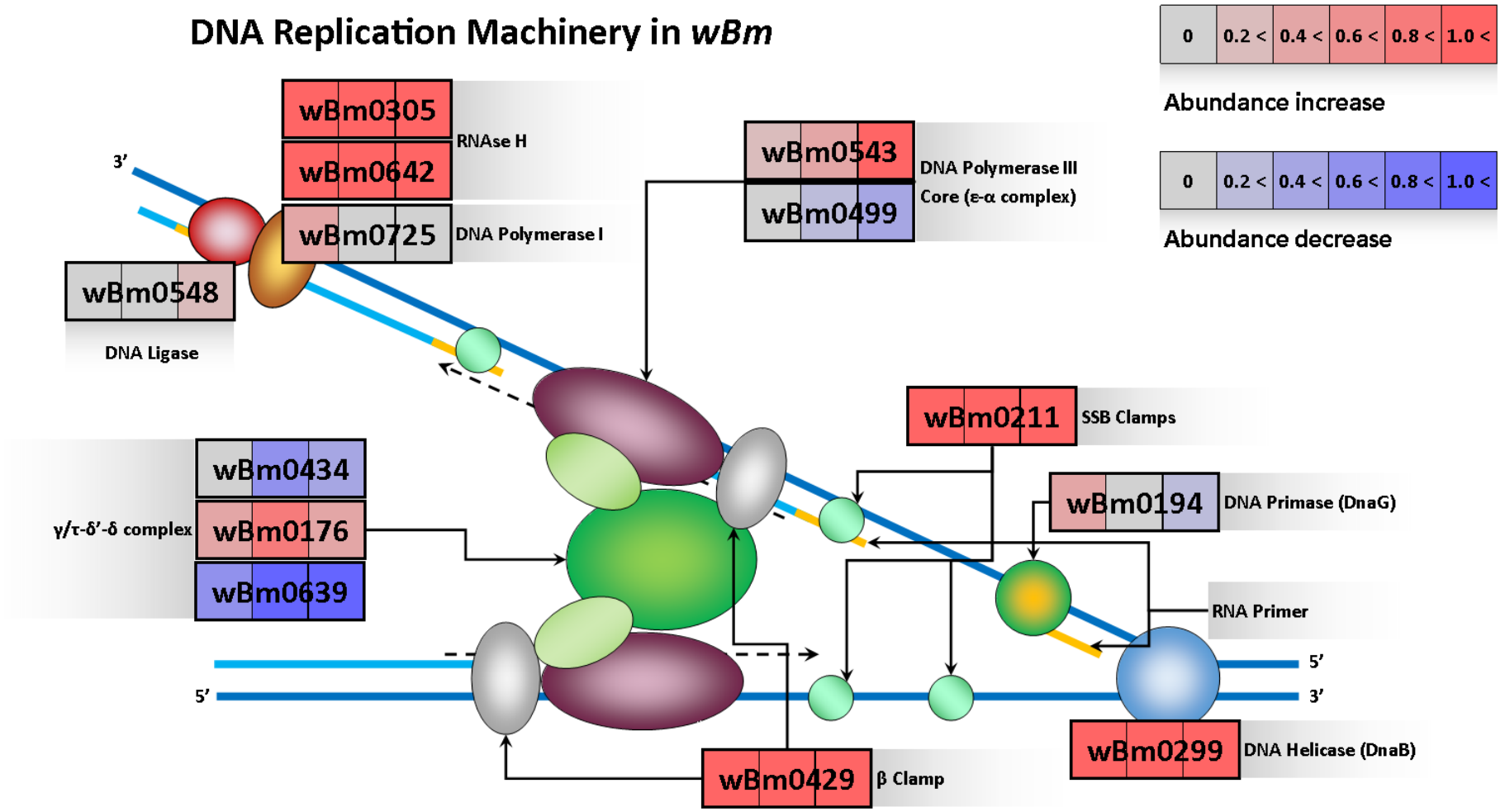


***Figure 3-8*: Illustration of the entirety of wBm's DNA replication machinery.** *Note that many of the clamp, helicase, and RNAse components of the pathway are upregulated across all studied time-point, but the key DNA polymerase core, and γ/τ-δ'-δ complex show several components that are downregulated.*

3.2.9 *Wolbachia* Electron Transport complexes II and III indicate upregulation

The electron transport chain in many organisms is comprised of 4 different complexes (Complex I to IV) plus ubiquinone, and are canonically used to generate a proton gradient, which could then be used by the ATP synthase complex (Complex V) to generate ATP.

This study identifies multiple components of Complexes II and III as being consistently upregulated across all studied time-points- both of which play a role in generating a proton gradient for ATP. Interestingly, only Complexes II and III show upregulation during studied time-points where Day 3 is used as a base-line (**Figure 3-9**). Examples of certain genes that show statistical significance in this pathway include *wBm0624* of Complex I (consistent downregulation, statistical significance at day 14), and *wBm0774* of Complex III (consistent upregulation, statistical significance at day 14). Furthermore, Complex V itself shows downregulation across all studied time-points, indicating that *w*Bm may not be utilising the proton gradient generated by Complex II and III for ATP biosynthesis. One potential reason as to why this is occurring, is that the generation of a proton gradient allows for the extrusion of sodium ions (Na+) to the extracellular space via Na+/H+ antiporter machinery, of which *w*Bm maintains several such antiporters[111]. This combination of H+ or Na+ gradients could then be used by *w*Bm for the export, or perhaps import, of metabolites to/from its nematode host.

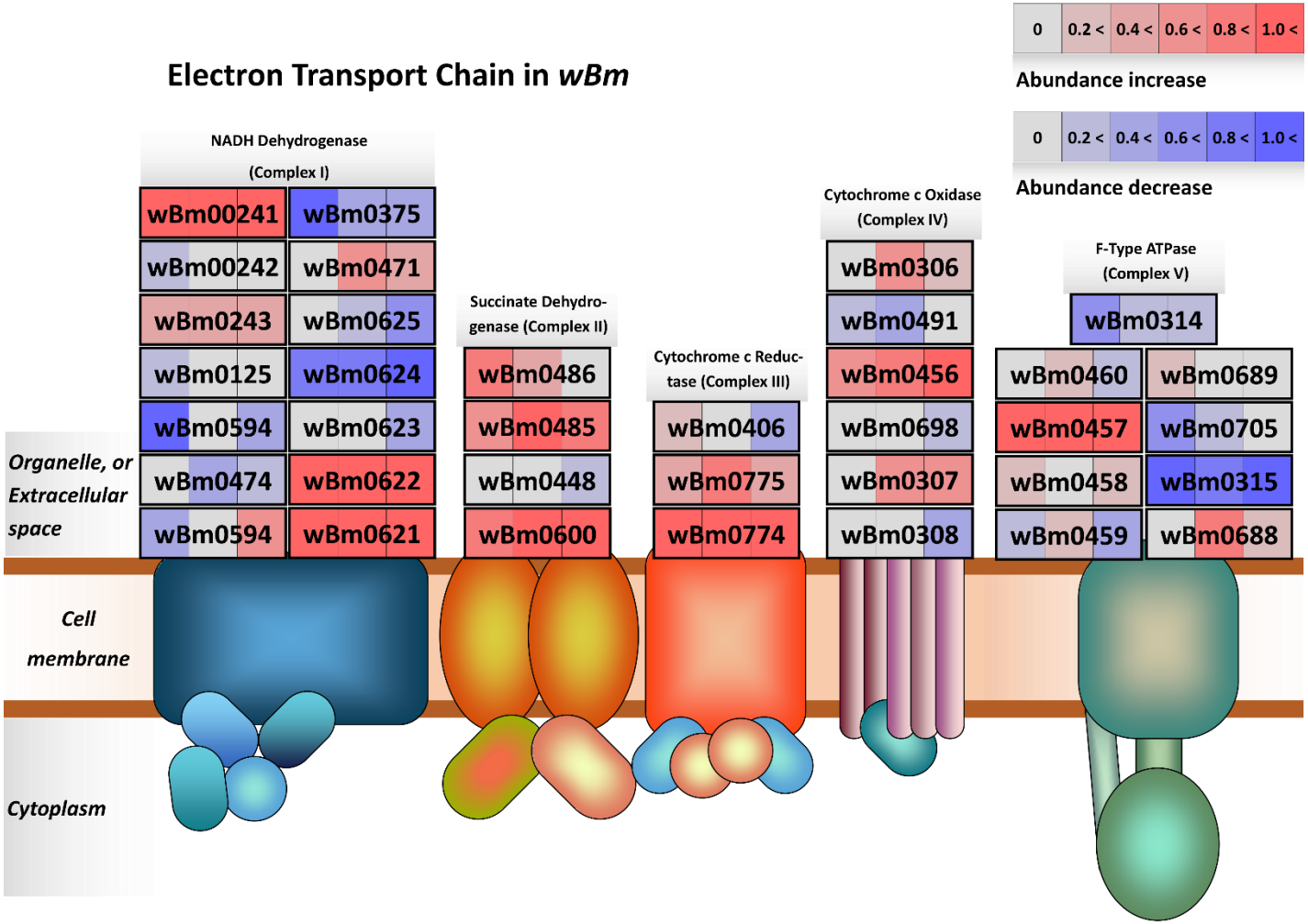


Figure 3-9: Illustration of the entirety of wBm's electron transport chain. Note how Complexes I, IV and V show a mixed transcriptomic profile, whilst Complexes II and III show upregulation.

3.2.10 *Wolbachia* transcriptome indicates upregulation in all components of haem biosynthesis

Previous studies have focused on how *w*Bm maintains the full pathway for haem biosynthesis, and how this contrasts with its host which lacks an equivalent pathway[111,247]. Instead, the nematode maintains a ferrocheletase enzyme of possibly bacterial origin, used in the final step of haem biosynthesis[265]. Haem is known to function as an essential cofactor for the proper function of many proteins critical for key biological processes, such as reproduction pathways[247].

This study has found that all 7 genes that make up the haem biosynthetic pathway in *w*Bm were upregulated across all studied time-points (**Table 3-3, Figure 3-10A**). This is further corroborated by genes encoding HemC (*wBm0777*) and HemE, (*wBm0001*) which showed statistically significant upregulation with EdgeR analysis (FDR < 0.05) during at least one pairwise comparison of this study (**Appendix 2**).

3.2.11 *Wolbachia's* Type IV secretion system indicates upregulation

Past studies have focused on *w*Bm's ability to interact with its nematode host and the potential involvement of *w*Bm's type IV secretion system in the symbiotic relationship via nutrient provisioning[111,159]. Canonically, the type IV secretion system is made up of 12 subunits: VirD4, and VirB1 through to VirB11, although in *w*Bm's case, it maintains genes that encode the 8 subunits VirD4, VirB3-VirB4, VirB6, and VirB8-11.

In this study, the data shows significant differential expression of two type IV secretion system components by EdgeR analysis during later time-point comparisons *wBm0793*, *wBm0282*, FDR < 0.05, (**Figure 3-10B, Appendix 2**). A further 3 of these 8 genes were also seen as upregulated across all time-points (*wBm0279*, *wBm0281*, *wBm0283*).

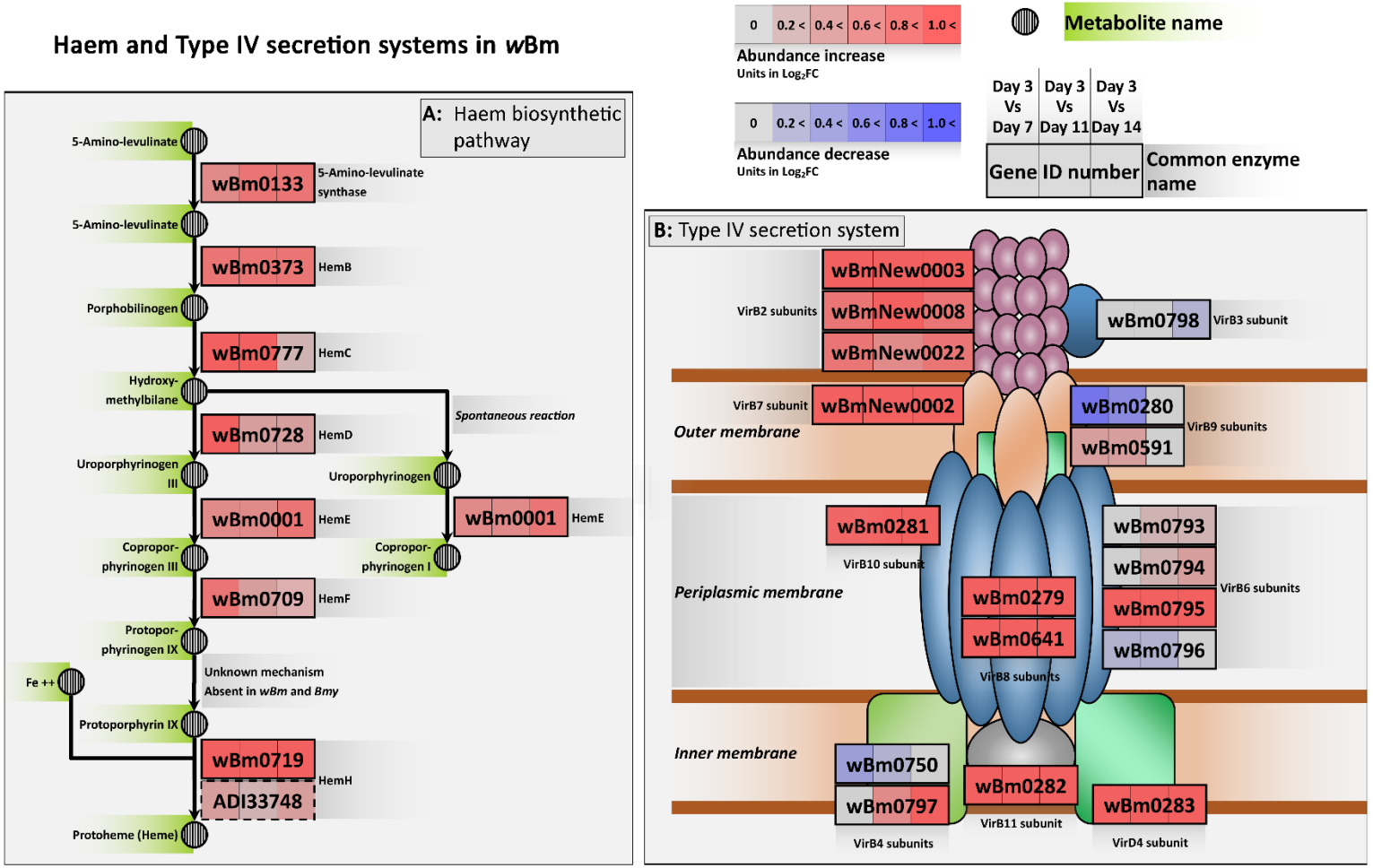


Figure 3-10: (A) Overview of the haem biosynthetic pathway in wBm, with the single gene in B. malayi also included (AD133748). (B) Structural overview of the Type IV secretion system in wBm, with newly identified genes in Chapter 2 included. *In (A), note the upregulation the entire pathway, aside from the gene that converts protoporphyrin IX to protoporphyrin IIX, which has shown to be absent in bacterial species which can produce haem. In (B), note how many of the core subunits are highly upregulated, including the 4 newly identified subunits (**Chapter 2**).*

3.2.12 Comparison of L3-L4 *w*Bm transcriptome with microfilarial transcriptome

Whilst the datasets are not directly comparable due to different experimental design, we attempted to identify any stage-specific transcripts unique to the L3-L4 stage *w*Bm transcriptome when compared to microfilariae-stage *w*Bm transcriptome (generated in **Chapter 5**). This was done to identify additional genes or signalling mechanisms that play a role in *w*Bm population growth, as these two time-points show markedly different behaviour; an actively replicating population during the L3-L4 stages, and a static population during the microfilariae stages. This comparison was achieved by identifying genes which show statistically significant differential expression between any of the L3-L4 stages and microfilariae stage, before then identifying genes that have a read count of 0 in either all three replicates in all time-points of the L3-L4 dataset, or the microfilariae dataset.

A total of 8 *w*Bm genes were shown to be expressed in all L3-L4 stages, but showed no expression in microfilariae stages. These include the SecF translocase subunit (*wBm0186*), phosphate ABC transporter (*wBm0231*), tRNA modification factor (*wBm0364*), nitropropane dioxygenase (*wBm0260*), 2 ribosomal proteins (*wBm0328*, *wBm0502*), and 2 hypothetical genes (*wBm0249*, *wBm0368*).

Conversely, there were no genes that showed stage-specific expression in microfilariae stage *w*Bm. This indicates that *w*Bm does not undergo any specific processes that are unique to the microfilariae stage.

3.2.13 Upregulated pathways of *Brugia malayi*

Similar to the methods initially used to analyse *w*Bm pathways, pathway enrichment techniques were applied onto the *B. malayi* transcriptome using the GAGE program[134,135] to identify differentially expressed pathways. As input, we utilised all genes that were identified as statistically significantly expressed during our consensus approach (**Figure 3-5**) with pathways reported if they have an FDR < 0.1 (a standard threshold for similar programs, such as Gene Set Enrichment Analysis[134]). Due to the biological complexity of *B. malayi* as compared to *w*Bm, differentially expressed pathways were more likely to be observed, and an automated solution would be necessary to investigate each pathway. As mentioned previously, investigating whole pathways will add additional biological context for the differentially expressed genes identified in *B. malayi* during the studied time-points.

Day 3 versus Day 7		Day 3 versus Day 11		Day 3 versus Day 14	
Pathway name (ID)	FDR	Pathway name (ID)	FDR	Pathway name (ID)	FDR
DNA replication (*bmy03030*)	2.61E-02	**bmy03460 Fanconi anaemia pathway**	1.78E-02	**DNA replication (*bmy03030*)**	8.38E-04

		DNA replication (*bmy03030*)	1.78E-02	**bmy03460 Fanconi anaemia pathway**	1.02E-02
				Biosynthesis of amino acids (*bmy01230*)	3.66E-02
				Homologous recombination (*bmy03440*)	7.54E-02
				Endocytosis (*bmy04144*)	7.54E-02
				FoxO signalling pathway (*bmy04068*)	7.54E-02
				Autophagy - animal (*bmy04140*)	7.54E-02
				mTOR signalling pathway (*bmy04150*)	3.43E-02
				ErbB signalling pathway (*bmy04012*)	8.79E-02
				Spliceosome (*bmy03040*)	8.79E-02

Table 3-4: ***Results obtained from GAGE showing pathways that are statistically significantly upregulated.*** *Note how the number of pathways identified increases as time progresses.*

The number of statistically significant pathways that show upregulation ranges from 1 to 10, with the number of identified pathways increasing as nematode development progresses (**Table 3-4**). Interestingly, the DNA replication pathway is the only pathway that was identified across all 4 time-points as being upregulated, likely linked to growth.

Additionally, during the Day 3 versus Day 14 time-point comparison, the autophagy pathway (*bmy04140*) was significantly upregulated (FDR < 0.1, **Table 3-4**, **Figure 3-11**). Whilst the upregulation of this pathway corroborates previous observations that the nematode regulates *Wolbachia* populations through activation of the autophagy pathway[266], the presence of several signalling pathways that are known to affect the autophagy pathway was also noted. Such pathways included the mTOR signalling pathway (*bmy04150*, **Table 3-4**), where upregulation of this pathway is known to both encourage cellular proliferation, but also suppresses autophagy via the action of the mTORC1 complex[267,268].

The mTOR pathway's upregulation in later time-points does instead correlate with *B. malayi*'s growth after the L3 to L4 moult. This is accompanied by the FoxO and ErbB signalling pathways (*bmy04068*, *bmy04012*), which have roles in cell proliferation and cell fate determination, likely related to the nematode's growth during the L4 stages.

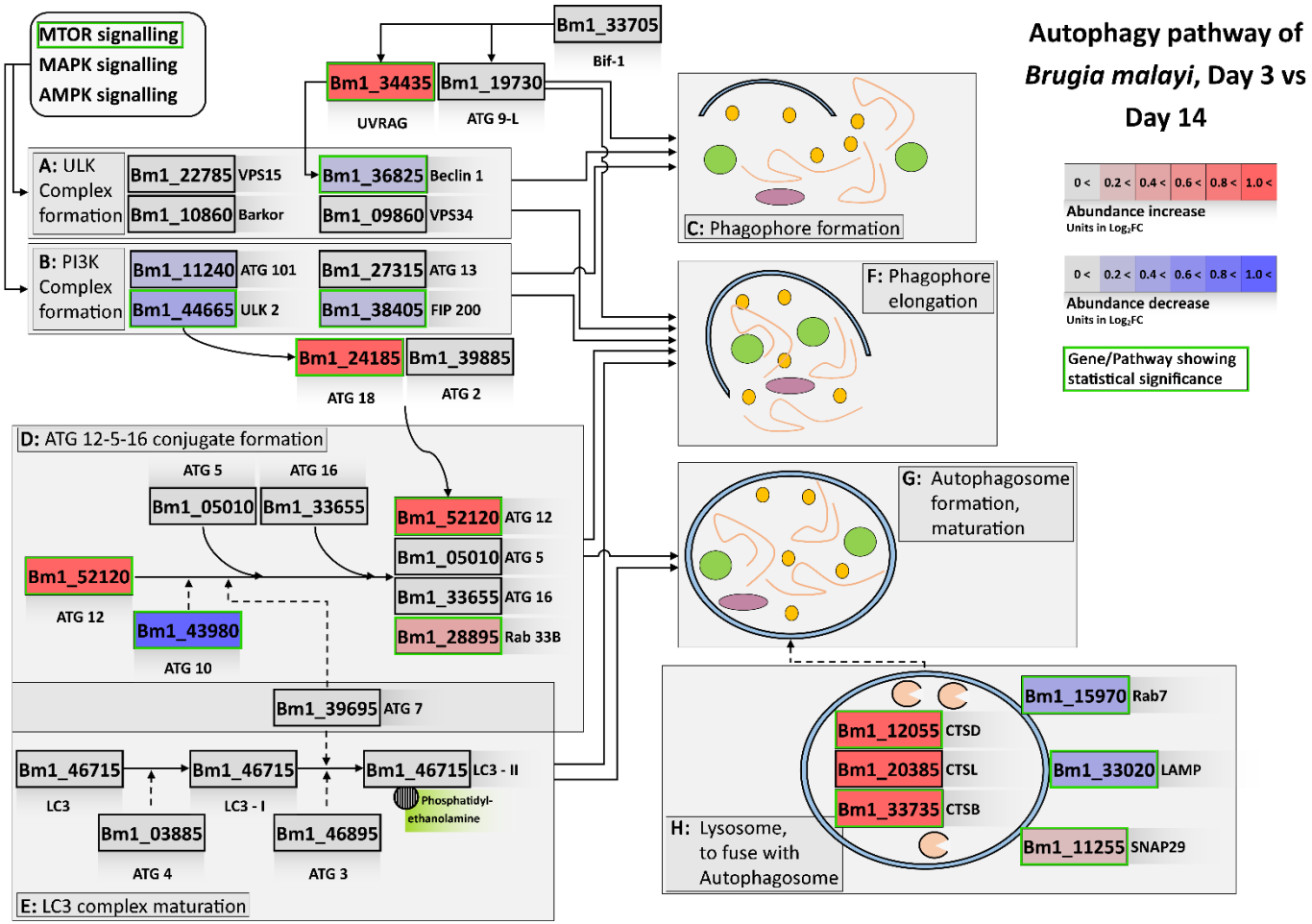


Figure 3-11: **Representation of genes involved in the autophagy pathway of *B. malayi*.** *Note the top left of the diagram, where there is a list of signalling pathways that are related to the initiation of this process that are not represented here.*

3.2.14 Downregulated pathways of *Brugia malayi*

A maximum of 10 pathways were identified as being downregulated during early time-point comparisons (**Table 3-5**). As time progresses and the infection continues, it is interesting to note that the number of downregulated pathways decreases, from 10 during early time-point comparisons down to 4 in the last time-point comparison (**Table 3-5**). This reduction of pathways observed to be downregulated may be related to the nematode moult from L3 to L4, as it has been observed in the past that nematode developmental moults are accompanied by a period of low metabolic activity[269,270].

Day 3 versus Day 7		Day 3 versus Day 11		Day 3 versus Day 14	
Pathway name (ID)	FDR	Pathway name (ID)	FDR	Pathway name (ID)	FDR
Ribosome (*bmy03010*)	3.46E-10	**Ribosome (*bmy03010*)**	3.79E-09	**Ribosome (*bmy03010*)**	5.44E-07
Oxidative phosphorylation (*bmy00190*)	3.40E-06	**Oxidative phosphorylation (*bmy00190*)**	1.75E-08	**Oxidative phosphorylation (*bmy00190*)**	5.44E-07
Metabolic pathways (*bmy01100*)	0.00023	**Metabolic pathways (*bmy01100*)**	6.20E-05	**RNA transport (*bmy03013*)**	4.63E-03
RNA transport (*bmy03013*)	0.00194	**Protein processing in endoplasmic reticulum (*bmy04141*)**	2.79E-04	**Metabolic pathways (*bmy01100*)**	4.63E-03
Spliceosome (*bmy03040*)	0.03425	**Spliceosome (*bmy03040*)**	3.82E-04		
Ribosome biogenesis in eukaryotes (*bmy03008*)	0.04805	**Endocytosis (*bmy04144*)**	2.57E-02		
Proteasome (*bmy03050*)	0.04805	**Ribosome biogenesis in eukaryotes (*bmy03008*)**	4.48E-02		
Valine, leucine and isoleucine degradation (*bmy00280*)	0.04805	**RNA transport (*bmy03013*)**	4.88E-02		
mRNA surveillance pathway (*bmy03015*)	0.08527	**Proteasome (*bmy03050*)**	6.12E-02		
Lysine degradation (*bmy00310*)	0.08527	**Glycan biosynthesis (*bmy00510*)**	8.11E-02		

***Table 3-5**: **Results obtained from GAGE showing pathways that are statistically significantly downregulated.** Note how the number of pathways identified decreases as time progresses, the opposite of pathways identified as upregulated.*

One interesting observation of this pathway analysis method is the presence of 4 pathways which show downregulation throughout all 4 time-points: The oxidative phosphorylation pathway (*bmy00190*), ribosome complex (*bmy03010*), RNA transport pathway (*bmy03013*), the and Metabolic pathways (*bmy01100*) (**Table 3-5**). Of these 4, the first pathway relates to energy production, whilst the next 3 pathways mentioned are related in some way to RNA and protein processing/maturation, although the term 'Metabolic pathways' encompasses many different pathways as well. It is perhaps interesting to speculate that the nematode may be reliant on the *w*Bm endosymbiont for much of its RNA or protein processing needs through the studied stages of development, alongside other pathways as mentioned previously.

3.2.15 Gene Ontology enrichment analysis of upregulated genes in *Brugia malayi's* L3-L4 transcriptome

Additional analysis on the *B. malayi* transcriptome was performed using Gene Ontology (GO) analysis via Blast2GO's Enrichment Analysis function[182], as described previously for the *w*Bm transcriptome. The background gene set was defined as the genetic content of *B. malayi*, with user-defined gene sets based genes that showed statistically significant differential expression via the consensus approach (**Figure 3-5**), and further split into upregulated or downregulated gene sets based on a $\log_2$ fold-change of > 0.2 or < -0.2 respectively- the minimum value that differentially expressed genes were identified. For analysis of the enriched GO terms, an FDR cut-off of < 0.05 was established as per default program settings to identify enriched GO terms. Following this, GO terms were then reduced to their most specific level via Blast2GO[182]. The full, un-summarised results of this analysis is included within **Appendix 2**.

13 GO terms were identified as enriched within the upregulated gene set of the Day 3 versus Day 7 comparison, after reducing terms to their most specific. These terms involved the nematode cuticle and moulting cycle, as well as integral membrane components and striated thin muscle filaments (**Table 3-6**). These terms are likely related to the nematode preparing a new cuticle for the developmental moult known to occur between days 8 and 11 after initial infection.

A total of 33 GO terms were identified as enriched within the upregulated gene set of the Day 3 versus Day 11 comparison after reduction. Terms identified in this comparison involve growth signalling and DNA unwinding, repair, and replication initiation (**Table 3-6**). These terms are likely related to nematode growth. Additional terms include ATP binding, as well as actin or muscle filaments (**Table 3-6**). The former term is likely related to the nematode's nucleotide salvage pathways, whilst the latter may be related to both the nematode's own growth, as well as *w*Bm's migration through *B. malayi* tissues, known to be mediated by actin filaments[152,271].

A final total of 44 GO terms were identified as enriched within the upregulated gene set of the Day 3 versus Day 14 comparison after reduction. Many of the GO terms identified in this set relate to growth and development, as well as sexual differentiation (**Table 3-6**). Examples of such terms include growth signalling, DNA replication initiation, cell division, and reproductive processes. In addition, we also see terms related to ATP binding and actin or muscle filaments, similar to Day 3 versus Day 11.

Day 3 versus Day 7 comparison (Total 13 terms)			
GO Category	GO ID	GO Name	FDR Value
Cellular Component	GO:0016021	Integral component of membrane	2.41E-05
Biological Process	GO:0018996	Moulting cycle, collagen and cuticulin-based cuticle	5.30E-04
Molecular Function	GO:0042302	Structural constituent of cuticle	5.30E-04
Cellular Component	GO:0005581	Collagen trimer	9.88E-04
Cellular Component	GO:0030017	Sarcomere	0.003263
Biological Process	GO:0006270	DNA replication initiation	0.004391
Cellular Component	GO:0042555	MCM complex	0.004391
Molecular Function	GO:0004857	Enzyme inhibitor activity	0.042115
Biological Process	GO:0051346	Negative regulation of hydrolase activity	0.044657
Biological Process	GO:0007264	Small GTPase mediated signal transduction	0.044657
Biological Process	GO:0045138	Nematode male tail tip morphogenesis	0.044657
Cellular Component	GO:0043235	Receptor complex	0.044657
Cellular Component	GO:0036379	Myofilament	0.044657
Day 3 versus Day 11 comparison (Total 33 terms)			
GO Category	GO ID	GO Name	FDR Value
Molecular Function	GO:0005524	ATP binding	9.36E-07
Cellular Component	GO:0005604	Basement membrane	1.71E-05
Biological Process	GO:0030239	Myofibril assembly	8.70E-05
Biological Process	GO:0008543	Fibroblast growth factor receptor signalling pathway	1.67E-04
Biological Process	GO:0032508	DNA duplex unwinding	3.81E-04
Molecular Function	GO:0003678	DNA helicase activity	3.81E-04
Biological Process	GO:0006270	DNA replication initiation	6.67E-04
Cellular Component	GO:0042555	MCM complex	6.95E-04

Cellular Component	GO:0031430	M band	7.64E-04
Cellular Component	GO:0005861	Troponin complex	0.003458
Cellular Component	GO:0032982	Myosin filament	0.003458
Molecular Function	GO:0008094	DNA-dependent ATPase activity	0.004718
Biological Process	GO:0031581	Hemidesmosome assembly	0.004894
Biological Process	GO:0006281	DNA repair	0.005126
Cellular Component	GO:0005581	Collagen trimer	0.008278
Biological Process	GO:0007018	Microtubule-based movement	0.012831
Biological Process	GO:0007155	Cell adhesion	0.014545
Cellular Component	GO:0099513	Polymeric cytoskeletal fibre	0.014545
Biological Process	GO:0048644	Muscle organ morphogenesis	0.016902
Cellular Component	GO:0031674	I band	0.016902
Biological Process	GO:0005977	Glycogen metabolic process	0.019072
Molecular Function	GO:0051015	Actin filament binding	0.019072
Biological Process	GO:0007528	Neuromuscular junction development	0.021953
Cellular Component	GO:0030054	Cell junction	0.021953
Molecular Function	GO:0016773	Phosphotransferase activity, alcohol group as acceptor	0.022933
Molecular Function	GO:0016301	Kinase activity	0.026598
Molecular Function	GO:0008017	Microtubule binding	0.027915
Biological Process	GO:0006310	DNA recombination	0.031129
Biological Process	GO:0006937	Regulation of muscle contraction	0.035291
Molecular Function	GO:0005200	Structural constituent of cytoskeleton	0.036329
Cellular Component	GO:0016010	Dystrophin-associated glycoprotein complex	0.044227
Molecular Function	GO:0003777	Microtubule motor activity	0.047571

Day 3 versus Day 14 comparison (Total 44 terms)			
GO Category	GO ID	GO Name	FDR Value
Biological Process	GO:0031581	Hemidesmosome assembly	2.13E-05
Molecular Function	GO:0005524	ATP binding	5.24E-05
Biological Process	GO:0006928	Movement of cell or subcellular component	1.08E-04
Biological Process	GO:0006281	DNA repair	1.23E-04
Biological Process	GO:0008543	Fibroblast growth factor receptor signalling pathway	1.39E-04
Biological Process	GO:0005977	Glycogen metabolic process	4.48E-04
Biological Process	GO:0006270	DNA replication initiation	7.31E-04
Cellular Component	GO:0042555	MCM complex	7.33E-04
Biological Process	GO:0030239	Myofibril assembly	0.001427
Molecular Function	GO:0019899	Enzyme binding	0.001989
Molecular Function	GO:0003678	DNA helicase activity	0.003012
Cellular Component	GO:0005861	Troponin complex	0.003622
Cellular Component	GO:0005604	Basement membrane	0.005827
Biological Process	GO:0006937	Regulation of muscle contraction	0.009313
Biological Process	GO:0051784	Negative regulation of nuclear division	0.010463
Biological Process	GO:0018107	Peptidyl-threonine phosphorylation	0.010463
Biological Process	GO:0040026	Positive regulation of vulval development	0.012445
Biological Process	GO:0001556	Oocyte maturation	0.012608
Biological Process	GO:0032508	DNA duplex unwinding	0.013659
Biological Process	GO:0031344	Regulation of cell projection organization	0.0151
Biological Process	GO:0048644	Muscle organ morphogenesis	0.016032
Cellular Component	GO:0015630	Microtubule cytoskeleton	0.017516
Molecular Function	GO:0004672	Protein kinase activity	0.018583

Biological Process	GO:0006323	DNA packaging	0.01945
Molecular Function	GO:0004373	Glycogen (starch) synthase activity	0.019773
Cellular Component	GO:0045120	Pronucleus	0.019773
Cellular Component	GO:0016459	Myosin complex	0.020907
Molecular Function	GO:0051015	Actin filament binding	0.021264
Biological Process	GO:0007229	Integrin-mediated signalling pathway	0.021653
Cellular Component	GO:0005884	Actin filament	0.021653
Cellular Component	GO:0031672	A band	0.027073
Biological Process	GO:0048812	Neuron projection morphogenesis	0.033496
Molecular Function	GO:0005102	Receptor binding	0.035917
Biological Process	GO:0009790	Embryo development	0.03644
Biological Process	GO:1905819	Negative regulation of chromosome separation	0.03644
Biological Process	GO:1902100	Negative regulation of metaphase/anaphase transition of cell cycle	0.03644
Biological Process	GO:0071173	Spindle assembly checkpoint	0.03644
Molecular Function	GO:0005200	Structural constituent of cytoskeleton	0.03644
Cellular Component	GO:0000776	Kinetochore	0.03644
Cellular Component	GO:0008305	Integrin complex	0.03644
Cellular Component	GO:0030055	Cell-substrate junction	0.03644
Biological Process	GO:0048667	Cell morphogenesis involved in neuron differentiation	0.043585
Biological Process	GO:0034728	Nucleosome organization	0.043985
Molecular Function	GO:0003777	Microtubule motor activity	0.048556

Table 3-6: ***List of GO terms enriched within the upregulated gene set of*** **B. malayi*****'s transcriptome during L3 to L4 development.*** *Note how the number of terms increases as time and development progresses, as well as how many of the terms identified in later time-points are related to growth signalling, cellular proliferation or differentiation, and actin/myosin filaments.*

3.2.16 Gene Ontology enrichment analysis of downregulated genes in *Brugia malayi's* L3-L4 transcriptome

A total of 20 GO terms were identified as enriched from the set of downregulated genes during the Day 3 versus Day 7 comparison after reduction (**Table 3-7**). Terms identified during this comparison involve reproductive processes, ATP synthesis and binding, cellular respiration, reproductive processes, ribosomal subunit assembly and translation of mRNA into proteins.

27 GO terms were identified in the downregulated gene set during the Day 3 versus Day 11 comparison after reduction. These terms involve the mitochondrial electron transport and respiratory chain, embryo development and reproductive processes, and moulting cycle (**Table 3-7**). Of note, there are a number of recurrent terms from the Day 3 versus Day 7 comparison that revolve around ribosomal subunit assembly and translation of mRNA into proteins.

The number of GO terms identified within the downregulated gene set continues to reduce to a low of 14 terms during the Day 3 versus Day 14 comparison after reduction. This set of terms is dominated by ribosomal subunits, their biosynthesis, and subsequent assembly (**Table 3-7**). Further terms also surround the mitochondrial ATP synthesis machinery as well. It is interesting to note that the enriched terms in all downregulated gene sets include terms related to ribosomal biogenesis and ATP biosynthesis, which overlaps with findings from previous pathway analysis by GAGE[135] (**Table 3-5**).

Day 3 versus Day 7 comparison (Total 20 terms)			
GO Category	GO ID	GO Name	FDR Value
Molecular Function	GO:0003735	Structural constituent of ribosome	9.55E-14
Biological Process	GO:0006412	Translation	1.11E-10
Molecular Function	GO:0016491	Oxidoreductase activity	5.92E-05
Molecular Function	GO:0015078	Hydrogen ion transmembrane transporter activity	4.18E-04
Biological Process	GO:0015986	ATP synthesis coupled proton transport	0.00101
Biological Process	GO:0042255	Ribosome assembly	0.001906
Cellular Component	GO:0022625	Cytosolic large ribosomal subunit	0.002952
Biological Process	GO:0007156	Homophilic cell adhesion via plasma membrane adhesion molecules	0.003294
Molecular Function	GO:0003756	Protein disulphide isomerase activity	0.003904
Biological Process	GO:0055114	Oxidation-reduction process	0.005878
Biological Process	GO:0008380	RNA splicing	0.011079
Cellular Component	GO:0005686	U2 snRNP	0.02173
Cellular Component	GO:0071013	Catalytic step 2 spliceosome	0.031303
Biological Process	GO:0048598	Embryonic morphogenesis	0.032946
Cellular Component	GO:0015935	Small ribosomal subunit	0.038633
Biological Process	GO:0006397	mRNA processing	0.038745
Biological Process	GO:0006338	Chromatin remodelling	0.039644
Molecular Function	GO:0005509	Calcium ion binding	0.039644
Biological Process	GO:0016192	Vesicle-mediated transport	0.04367
Cellular Component	GO:0000276	Mitochondrial proton-transporting ATP synthase complex, coupling factor F(o)	0.04367

Day 3 versus Day 11 comparison (Total 27 terms)			
GO Category	GO ID	GO Name	FDR Value
Cellular Component	GO:0005840	Ribosome	6.72E-06
Cellular Component	GO:0016021	Integral component of membrane	4.35E-05
Molecular Function	GO:0003735	Structural constituent of ribosome	2.59E-04
Cellular Component	GO:0071013	Catalytic step 2 spliceosome	4.29E-04
Biological Process	GO:0015986	ATP synthesis coupled proton transport	9.65E-04
Cellular Component	GO:0005783	Endoplasmic reticulum	0.001572
Biological Process	GO:0000398	mRNA splicing, via spliceosome	0.001735
Biological Process	GO:0006412	Translation	0.002111
Biological Process	GO:0015991	ATP hydrolysis coupled proton transport	0.00297
Cellular Component	GO:0000276	Mitochondrial proton-transporting ATP synthase complex, coupling factor F(o)	0.003176
Molecular Function	GO:0003756	Protein disulphide isomerase activity	0.003647
Molecular Function	GO:0004298	Threonine-type endopeptidase activity	0.004749
Cellular Component	GO:0005839	Proteasome core complex	0.004749
Biological Process	GO:0042335	Cuticle development	0.005218
Molecular Function	GO:0008137	NADH dehydrogenase (ubiquinone) activity	0.00537
Molecular Function	GO:0046961	Proton-transporting ATPase activity, rotational mechanism	0.005642
Biological Process	GO:0018996	Moulting cycle, collagen and cuticulin-based cuticle	0.008914
Biological Process	GO:0022618	Ribonucleoprotein complex assembly	0.011034
Cellular Component	GO:0005685	U1 snRNP	0.019459
Biological Process	GO:0019915	Lipid storage	0.023043
Cellular Component	GO:0033176	Proton-transporting V-type ATPase complex	0.025148
Biological Process	GO:0006898	Receptor-mediated endocytosis	0.029803
Biological Process	GO:0000003	Reproduction	0.030552

Cellular Component	GO:0071010	Pre-spliceosome	0.032016
Biological Process	GO:0042775	Mitochondrial ATP synthesis coupled electron transport	0.040036
Molecular Function	GO:0019843	rRNA binding	0.040152
Biological Process	GO:0006457	Protein folding	0.041382
Day 3 versus Day 14 comparison (Total 14 terms)			
GO Category	GO ID	GO Name	FDR Value
Molecular Function	GO:0003735	Structural constituent of ribosome	2.68E-05
Biological Process	GO:0006412	Translation	2.70E-05
Cellular Component	GO:0016021	Integral component of membrane	9.60E-05
Molecular Function	GO:0005509	Calcium ion binding	5.30E-04
Biological Process	GO:0015986	ATP synthesis coupled proton transport	0.002015
Molecular Function	GO:0015078	Hydrogen ion transmembrane transporter activity	0.002686
Cellular Component	GO:0000276	Mitochondrial proton-transporting ATP synthase complex, coupling factor F(o)	0.00599
Biological Process	GO:0000028	Ribosomal small subunit assembly	0.006716
Cellular Component	GO:0022627	Cytosolic small ribosomal subunit	0.013313
Biological Process	GO:0030322	Stabilization of membrane potential	0.030124
Molecular Function	GO:0022841	Potassium ion leak channel activity	0.030124
Biological Process	GO:0044707	Single-multicellular organism process	0.035489
Biological Process	GO:0048856	Anatomical structure development	0.041754
Biological Process	GO:0044767	Single-organism developmental process	0.045174

Table 3-7: Lists of GO terms identified as enriched within the downregulated gene set of **B. malayi***'s transcriptome across the 4 time-points studied.*

3.3 Discussion

Here we illustrate the first study to comprehensively investigate the developmental transcriptome of *B. malayi* and its *w*Bm endosymbiont during the first two weeks of infection of the mammalian host. To the authors' knowledge, this study provides the first set of experimental evidence to investigate previous hypotheses in how *w*Bm may act as a major source of nucleotides and haem for the nematode from as early as the L3-L4 developmental stages. In addition to this investigation, this study also corroberates several new pathways that may play roles in the endosymbiotic relationship, and potentially fuel *w*Bm's rapid growth that is known to occur during the studied period.

It should be noted that all samples from Day 3 had comparatively low read counts, which will have knock-on effects on predictions of statistically significant differentially expressed genes and increases the possibility of false-positive results from RNA-sequencing work. Due to the nature of *w*Bm as an obligate endosymbiont, such low read counts from RNA-sequencing is an unavoidable problem that needs to be taken into account. As a result, further technical validation would greatly improve and strengthen the results of this analysis, particularly with the predictions made from the glycolysis-TCA cycle. This could take the form of qPCR experiments to corroborate expression data from RNA-seq, which is covered in additional detail in **Chapter 4**.

Regardless of this, the results observed identify several pathways of interest that are supportive of the literature currently available. First of which is the alternating transcriptional profile observed within the glycolytic/gluconeogenic and TCA cycle pathways of *w*Bm, which has been observed in many other organisms ranging from cancer cells[272], to yeasts[273], to protozoan parasites such as malaria[274]. This effect is termed aerobic glycolysis, and in cancer cells the Warburg effect. Despite the wide range of organisms that utilise this process, a key theme is the upregulation of the glycolytic pathway, and downregulation of the TCA cycle even within oxygen-rich environments[272]. One of the main theories to explain why this effect takes place, despite being energy-inefficient, is that the rate of flux of the glycolytic pathway is significantly higher than that in canonical respiration pathways[272,273,275]. Thus, in nutrient rich environments where resource efficiency is not a priority, higher quantities of ATP can be generated using glycolysis as compared to oxidative phosphorylation. Another theory behind this process, is that the glycolytic pathway produces many of the essential intermediate metabolites required by cells to proliferate[272,274]. The situation in *w*Bm may differ, however, as certain key enzymes required to utilise glucose via the glycolytic pathway, such as

phosphoglucomutase or glucose phosphate isomerase, are absent[111] (**Figure 3-6**). Instead, *w*Bm maintains genes that can be used in the reverse, gluconeogenic, pathway, (e.g. Fructose-1-bisphosphatase, *wBm0132* and *wBm0158*, or pyruvate phosphate dikinase, *wBm0209*,[156]) leading to hypotheses that *w*Bm does not perform glycolysis, but instead relies on gluconeogenesis[23]. As such, gluconeogenesis starting from pyruvate, rather than glycolysis, may provide many of the key building blocks for cellular proliferation, although at an energetic cost. As to where pyruvate may be sourced from, previous studies have indicated that pyruvate may be actively supplied by the host to the *Wolbachia* endosymbiont, due to the observation of host glycolytic enzymes being localised to *Wolbachia* surface proteins[152,276], as well as observations of the deleterious effects nematode glycolysis inhibition by RNA interference has on *w*Bm populations[276]. Additionally, recent work by Voronin *et al.*[277] further emphasises the importance of pyruvate in *w*Bm maintenance within host tissues[277]. By chemically inhibiting glycolysis enzymes in adult nematodes, Voronin *et al.* observed both a reduction in *w*Bm populations in male nematodes , as well as a reduction in microfilaria released from adult females- a hallmark of *w*Bm depletion via antibiotics. These effects were also determined by Voronin *et al.* to be at least partially reversible via the addition of extracellular pyruvate[277].

These observations help to validate the role of pyruvate in nematode-*w*Bm symbiosis, and how the nematode may act as a key carbon source for *w*Bm. Such results, while made in adult nematodes, are potentially applicable to other life cycle stages of the nematode, such as the L3-L4 developmental stages as is predicted by the transcriptomics work from this study. This is because adult females are constantly producing microfilariae for release, each of which maintains a small population of *Wolbachia*[15,67]. As such, one could predict that the *w*Bm population in adult females is constantly undergoing a period of population expansion to maintain its numbers, and utilising similar pathways to support the rapid expansion of *w*Bm population as examined in this study.

Alongside the described relationship between the glycolysis-TCA cycle and *w*Bm population dynamics, the transcription profile of *w*Bm's DNA replication machinery also indicates a similar, if slightly delayed, trend (**Figure 3-8**). Specifically, multiple key components of the DNA polymerase III enzyme indicated downregulation from Day 11 onwards. These subunits make up the catalytic DNA polymerase subunit (subunits alpha, gamma/tau), as well as the DNA clamp 'loader' (delta-prime), which plays a role in DNA replication initiation complex formation, and efficient DNA replication[260,261]. This observation is corroborated by use of the time-course based analysis, which also show downregulation of these same components

from Day 11 onwards, as well as several additional enzymes involved in the process such as DNA polymerase I, and RNAse H I and II. Many of these additional enzymes involved in DNA replication also play a role in DNA repair[263,264,278], with these genes being upregulated across all time-points possibly for the purpose of maintenance, rather than active replication.

In addition to these observations surrounding *w*Bm population growth mechanisms, genes such as those related to surface proteins or transport mechanisms (Type IV) were noted to be highly upregulated (**Figure 3-10**), possibly indicating *w*Bm's focus on its own cellular proliferation (**Appendix 2**). Additional genes, such as those involved in haem and nucleotide biosynthesis pathways, have been previously hypothesised to be important for nematode host embryogenesis or development[111,112,247]. Observations on how these pathways are upregulated in *w*Bm lends further support the theory that *w*Bm provides these components to the nematode host at key points of metabolic demand. It is interesting to speculate that *w*Bm may be proactively maintaining a homeostatic level of key metabolites required for continued nematode growth. Upon depletion of *w*Bm from the system via antibiotics, these metabolites may be consumed by *Brugia malayi* during the moult, leaving little to continue development through the L4 stage and onwards. Such a theory will require an in-depth metabolomics study of these life-cycle stages in a treated and untreated group to confirm this.

When comparing the transcriptomes of *w*Bm between the microfilariae and L3-L4 developmental stages, no genes were identified as being microfilariae-stage specific. The lack of expression of *w*Bm stage-specific genes during the microfilariae stage indicates that the endosymbiont does not engage any specific pathway to aid the nematode host. This is in line with previous observations that show minimal deleterious effects on microfilariae viability should the *w*Bm endosymbiont be removed by antibiotics[85,279]. In comparison, 12 genes were identified as L3-L4 stage-specific, which include ribonuclease III and tRNA modification factors, as well as the SecF translocase. Genes that are responsible for mRNA turnover play an important role during protein biosynthesis and gene regulation (as reviewed by Garneau *et al.*[280]), whilst the SecF gene plays a part in the Sec translocase system for protein localisation and insertion into cellular membranes (as reviewed by du Plessis *et al.*[197]). The lack of expression in these genes during microfilariae-stage *w*Bm indicates reduced protein processing ability, and possibly host-endosymbiont signalling rate due to a lack of membrane-inserted proteins. While reduced protein processing ability can be expected in a static *w*Bm population, it is interesting to speculate that the lack of membrane-embedded

proteins may also play a role in *w*Bm population maintenance, via evasion of the host autophagy pathway.

Within the *B. malayi* host, genes and GO-terms related to the nematode cuticle were observed to be specifically upregulated at early time-points, with many signalling and biological pathways related to DNA replication, repair, cell fate, differentiation and proliferation upregulated during later time-points (**Table 3-6**). This correlates with previous literature that describes the nematode undergoing significant increases in size during the first few weeks of infection of the mammalian host[281], as well as development of sexual organs to become reproductively active adults. As the nematode likely lacks the biosynthetic pathways in full to fuel this development and growth, it must salvage resources such as nucleotides from an exogenous source, likely *w*Bm. This is shown by the upregulation of GO terms related to ATP binding, and the associated nucleotide/nucleoside binding terms (**Appendix 2**) in *B. malayi*, as well as the nucleotide biosynthetic pathways in *w*Bm that show upregulation across all time-points (**Figure 3-7**). This further strengthens previous predictions that nucleotide biosynthesis and provisioning play a central role in *Wolbachia*-nematode symbiosis, within both reproductively active adults and actively developing larval stages[111].

Separately, many upregulated pathways seen in this dataset that encourage cellular proliferation are known to have an antagonistic effect on the autophagy pathway[282–284]. The mTOR and MAPK signalling pathways in particular are known to be involved in suppressing the initiation of the autophagy process, specifically the formation of the ULK and PI3K complexes (**Figure 3-11**). Inhibition of these complexes prevents the formation of the phagophore and its eventual maturation into the autophagosome, thus inhibiting the autophagy pathway[282–284]. This runs counter to other observations made from this transcriptomic analysis (**Table 3-4**, **Figure 3-11**), which corroborates the role of the nematode autophagy pathway in controlling *w*Bm population, particularly during the L4 stages that were studied[266]. This implies that the nematode must maintain a careful balancing act between the need to encourage its own cellular proliferation, whilst simultaneously regulating the population of its *Wolbachia* endosymbiont during the first few weeks of infection of the mammalian host. The timing of the autophagy pathway's activation is coincident with the slowdown of *w*Bm population growth[15], implying that even when the nematode has evolved a mutualistic association with *Wolbachia*, it is still recognised by the host's immune system as a 'pathogen', and regulated to enable sufficient resource provisioning without incurring fitness costs. It is interesting to speculate that due to *w*Bm's localisation in an intracellular vesicle, the nematode may circumvent the need to

generate the initial phagophore via stimulation of the ULK and PI3K complexes. The nematode may instead directly mature the vesicles containing *w*Bm into autophagosomes, thus avoiding the potential suppressive effects these complexes may have on encouraging cellular proliferation and growth. To an extent, this can be observed in the transcriptome data of this study, where components of the ATG12-5-16 conjugate are specifically upregulated (**Figure 3-11**).

To conclude, the analysis of gene transcription during the L3 to L4 development of the nematode has provided several valuable insights into the biology underpinning the symbiotic relationship between *w*Bm and *Brugia malayi*. Expression studies corroborate several phenotypic observations and predictions made previously and shed light on the potential role of the Warburg/Crabtree effect in the bacterium's proliferation, as well as the potential role that pyruvate plays within the symbiotic relationship. These results will help to form a better understanding of the energy metabolism and interplay between endosymbiont and host, potentially informing rationale targets for drug discovery and development strategies.

3.4 Materials & Methods

3.4.1 Parasites

Mammalian-stage L3 to L4 *B. malayi* larvae, 3, 7, 11, and 14 days post infection (2,000 nematodes for days 3 and 7, 1,000 nematodes for days 11 and 14), were collected from intra-peritoneal infections of Mongolian jirds (*Meriones unguiculatus*) at the Filariasis Research Reagent Resource Centre. The laboratory isolate had been maintained by serial passage in jirds and cats since first derived from a human infection in Malaysia more than 20 years ago. In order to preserve the RNA, the in batches of 150-250 nematodes were collected into 150µl PBS and immediately frozen in liquid nitrogen and shipped to the Liverpool School of Tropical Medicine on dry ice. 10 nematode larvae were stored individually for *Wolbachia* load analysis by quantification PCR.

3.4.2 RNA extraction from *Brugia malayi* and *Wolbachia*

RNA was extracted from *B. malayi* and *Wolbachia* using the TRIzol Plus RNA extraction kit (Invitrogen), which used a modified protocol for RNA extraction to that given by the manufacturers.

Briefly, nematode batches were defrosted and simultaneously pelleted by centrifugation (Eppendorf) at 4,500rpm at 4°C for 10 minutes. The supernatant was removed and 200µl of TRIzol reagent was added to the samples. Batches for each life-cycle stage were pooled into sterile 2ml screw cap tubes containing ceramic beads of multiple sizes (1.4 - 2.8mm) (CKMix,

Bertin). Batches were homogenized at 6,000rpm (Minilys, Bertin Instruments) for 4 x 30 seconds, cooling on dry ice for 30 seconds in between. The homogenized sample was removed from the beads before being washed with a further 200µl of TRIzol to maximize sample recovery, before storage at -80°C overnight.

The samples in TRIzol were thawed and subsequently incubated at room temperature for 5 minutes. 200µl molecular grade Chloroform (Invitrogen) was added for every 1ml of TRIzol reagent, and mixed by vigorously shaking the tube, followed by incubation at room temperature for three minutes. The sample was then centrifuged at 12,000xg for 15 minutes at 4°C. The upper aqueous phase (approximately 600µl) was removed and transferred to a nuclease-free 1.5ml tube. An equal volume of ice cold 70% ethanol was added to the sample and vortexed for 2 - 3 seconds. The tube was inverted several times to disperse any precipitate generated from the addition of ethanol.

Subsequent binding, washing and elution of RNA from the sample was then conducted in accordance to manufacturer's instructions. Final RNA was eluted in 2 x 50µl molecular grade water (Invitrogen).

3.4.3 qPCR analysis of *w*Bm population

DNA was extracted from the individual nematode larvae using DNeasy Blood & Tissue kit (Qiagen) following manufacturer's instructions with modifications. Briefly, individual nematodes were thawed and placed in separate microcentrifuge tubes, before incubated in 180µl of Qiagen's Buffer ATL and 20µl of Qiagen's proteinase K, and incubated overnight at 56°C. Tubes were centrifuged, and 200µl of Qiagen's Buffer AL was added, vortexed for 15 seconds, and incubated at 70°C for 10 minutes. The samples were centrifuged, then 200µl of ethanol was added to the sample and vortexed for 15 seconds, before centrifuged again. Samples were then transferred into a DNeasy Mini spin column placed in a 2ml collection tube, and centrifuged at >6,000 x g for 1 minute, with flow-through discarded. The column was placed in a new collection tube, and 500µl of Qiagen's Buffer AW1 was added and centrifuged for 1 minute at 6,000 x g, with flow-through discarded. The column was placed in a new collection tube, and 500µl of Qiagen's Buffer AW2 was added, and the column centrifuged for 3 minutes, with flow-through discarded. The column was placed into a new microcentrifuge tube, and 40µl of buffer AE was added before the column was left to stand at room temperature for 1 minute. The column was centrifuged for 1 minute at 6,000 x g, before the amount of recovered DNA was quantified on a Nanodrop Spectrophotometer (Thermo).

Amplification of the *wsp* gene utilised a final concentration of 0.30µM Taq Probe/Primer. Amplification of the *gst* gene utilised a final concentration of 0.15 µM for each primer. All tests were carried out in quadruplicate, within a total reaction volume of 20µl and included 1µl of DNA from each sample, 1µl of *gst*-VIC probe at a final concentration of 0.25µM, and 10µl of Fast Advanced Master Mix (Thermo). For both *wsp* and *gst*, a standard curve was also generated in quadruplicate. All qPCR reactions were conducted using a CFX 384 Real Time PCR detection system (BioRad), with all reactions heated for an initial 2 minutes at 50°C and 20 seconds at 95°C, before subjected to 40 cycles of 3 seconds at 95°C, and 30 seconds at 60°C.

3.4.4 Illumina RNA Library Preparation and Sequencing

In order to remove excess large ribosomal RNA from both *Wolbachia* and *B. malayi*, total RNA was treated with Terminator 5'Phosphate-Dependent Exonuclease (Epicentre), in accordance to the manufacturer's instructions.

Illumina GA-IIx cDNA libraries were prepared using the ScriptSeq v2 library preparation kit (Epicentre). Two separate paired-end sequencing runs (2x100bp reads) were performed on an Illumina GA-IIx instrument at the Centre for Genomic Research (CGR), University of Liverpool (UK). One additional run using the same chemistry was performed with a similar instrument at GenePool, University of Edinburgh (UK). Read files for all data were converted from raw basecall to fastq using CASAVA 1.8 (Illumina). The raw Fastq data files were trimmed for the presence of Illumina-specific adapter sequences using Cutadapt[243] (v1.2.1) with the option–*O 3* and further trimmed for quality using Sickle[244] (ver.1.33). Quality scores were assessed using FASTQC[245] (v0.9.2).

3.4.5 Mapping and Quantifying RNA-sequence Data

A strategy to retain RNA-sequence reads via non-unique mapping was utilised, allowing for the preservation of data relating to lateral gene transfer events, which would potentially be lost when mapping to a concatenated genome. RNA-sequence reads from the processed fastq files were aligned to the genome of *Brugia malayi* via Subread-Aligner[188] (ver. 1.5.0), giving a Binary Alignment/Map (BAM) output file that contained aligned and unaligned sequences. The BAM file was separated into aligned and unaligned reads using Samtools[221] (ver. 0.1.19), with one BAM file containing aligned RNA-sequence reads mapped to *Brugia malayi*. The second BAM file that contained only unaligned reads were then remapped against the *w*Bm genome using Subread-Aligner, resulting in a final BAM file that contained

aligned and unaligned RNA-sequence data to *w*Bm. Aligned RNA-sequence data in this file was unique to *w*Bm only.

The aligned RNA-sequence data from both *B. malayi* and *w*Bm were quantified against their respective genome annotations held on the NCBI database (assembly ASM299v2 and ASM838v1 respectively). This quantification used the program FeatureCounts[252] (version 1.5.0-p3) to obtain read counts that were then subsequently used in differential expression analysis.

3.4.6 Differential Expression Analysis and Pathway mapping

Differential expression analysis was conducted on both the *B. malayi* and *w*Bm datasets using either the programs CuffDiff (version 2.2.1), DESeq2 (version 1.22.1), or EdgeR (version 3.10.5)[127–130]. The former utilised Fragments Per Kilobase per Million mapped reads (FPKM) quantified directly from the aligned .BAM files to a .General Feature Format (GFF) or General Transfer Format (GTF) file, whilst the latter two utilised read counts obtained from the program FeatureCounts[252], with EdgeR results being obtained via the online wrapper Degust (version 3.2.0)[259]. Pairwise comparisons for all possible combinations of time-points were conducted for both organisms, with a primary focus on the 'base-line' series of comparisons followed by the 'time-course' series of comparisons to act as additional validation of observations. Genes were deemed as statistically significantly differentially expressed using a FDR, of < 0.05 – a standard setting in all three utilised programs.

From this list of statistically significantly differentially expressed genes, gene ontology term enrichment was performed using the platform Blast2GO, via a single-tailed Fisher's Exact Test[182]. The datasets from *B. malayi* and *w*Bm were treated separately during this analysis, with gene ontology terms being deemed as statistically significant using a FDR of < 0.05.

For pathway analysis of the *w*Bm transcriptome, the fold-change results from EdgeR analysis was then input into the PathView program[258], ver. 1.22.0 using the options *kegg.native = T, node.sum = mean*. The fold-change results from EdgeR were also used as input for BioCyc's Pathway Collage program[255–257] to generate individual pathways for manual inspection.

For pathway analysis of the *B. malayi* transcriptome, gene IDs were converted using BioDB.net's db2db online resource[285], from GenBank nucleotide accession to KEGG gene ID. These converted gene IDs, and their associated fold-change results from EdgeR analysis were used as input into the PathView program[258] (ver. 1.22.0), using the options *kegg.native = T, node.sum = mean*. In addition, the fold-change results from EdgeR was also used as input for the Generally Applicable Gene set Enrichment program[135] (ver. 2.32.0), using the options

gsets = kegg.gs, same.dir = TRUE, saaTest = gs.KSTest for analysis of up- or down-regulated pathways.

Chapter 4 Identification of a new, translated *Wolbachia*-specific gene, and validation of RNA-sequencing results via quantification PCR

4.1 Introduction

RNA-sequencing is a powerful technique that has been utilised for a range of purposes in the years since its initial development, primarily for *de-novo* transcriptome construction of non-model organisms, or measuring gene expression. This allows identification of genes of interest in biological systems that may be differentially expressed under different conditions. Examples of this include analysis of expression across different life cycle stages of various pathogens, or analysis of gene expression upon exposure to experimental conditions, such as drug treatment, pathogen exposure, or environmental changes[174,286–288]. The ability to measure gene expression has been previously available in the form of Microarrays[123], however such arrays require prior knowledge in the form of either an annotated genome, or an assembled transcriptome, of the organism of interest to be available for generating the array[123]. By contrast, RNA-sequencing does not rely on the existence of a published genome, instead sequencing all reads within a sample directly. These sequences can then either be aligned to an existing genome for quantification of differential expression, or be used to generate a *de-novo* transcriptome from scratch, before quantification of differential expression[173,289].

In addition, RNA-sequencing is a powerful, but under-utilised, technique for aiding genome annotation projects, or identification of novel transcripts in previously sequenced genomes[173]. Most basic genome annotation tools scan genomes for the presence of open reading frames that are flanked by suitable start- and stop- codons, or possibly the genome as a whole is screened against existing databases to identify protein-coding genes[176,178,290]. However, this does not exploit the central dogma of molecular biology, which involves the two-step process of protein-coding DNA being transcribed into mRNA, before being translated into proteins. Thus, whilst detection of open reading frames within a genome may indicate the presence of coding regions of DNA, mRNA species that align to these open-reading frames significantly strengthens any predictions of a coding gene being present, and can simultaneously aid in identifying new, previously unidentified genes[291]. In addition, 'non-coding' or functional mRNAs are frequently present within genomes that do not abide by the rules of protein translation, and thus may be difficult to detect without measuring mRNA quantities directly.

For all of the benefits of RNA-sequencing however, due to sample constraints and/or the various processing steps required before RNA-sequencing data can be analysed, there is an inherent degree of uncertainty or noise with obtained results[173,289]. Results from RNA-sequencing data could therefore benefit greatly from alternative experimental methods to validate results, or generate more comprehensive hypotheses. Examples of such techniques include Reverse-Transcription (Quantification) Polymerase Chain Reaction (RT-(q)PCR), or additional proteomics experiments.

RT-qPCR relies on the ability to directly measure the amplification of small amounts of DNA from a starting concentration, as well as the ability for reverse transcriptase enzymes to synthesis complementary DNA (cDNA) sequences from template mRNAs. These PCR-based techniques are a mainstay of many molecular biology laboratories, and offer a quick, accurate, cost-effective, and reproducible method of measuring the expression of a small number of genes within a variety of different samples. Such measurements can be done either via absolute quantification, which utilises a standard curve to directly measure the number of sequences present (e.g. biomarkers), whereas relative quantification normalises to a reference gene[292,293]. Regardless of the method, the efficiency of the PCR reaction must first be determined to compare amplified samples and their starting copy numbers. PCR efficiency is a measure of how well the PCR product doubles with each thermocycle, and can vary based on the primer sequences and reaction mixture. Once optimised, RT-qPCR methods can provide reproducible measurements of a target gene's transcription profile at relatively low cost in many sample types.

By contrast, proteomics relies on inferring the presence of protein sequences via detection of their constituent peptides after digestion by a proteolytic enzyme. Subsequent to this digestion, the samples are separated by liquid chromatography to separate samples into constituent peptide fragments. Samples are then subjected to Mass Spectrometry (MS), which first ionises and aerosolises the peptide fragments, which can then be further fragmented in some way, such as collision-induced fragmentation using an inert noble gas (e.g. helium or argon). These fragments are then sent through a mass analyser, which separates the fragments based on their mass-to-charge ratios, which are in turn based on the physiochemical properties of the peptide's constituent amino acids. Detection of these mass-to-charge ratios can then be analysed computationally by programs such as Sequest[294] or Mascot[295] to identify the parent peptide fragment. These programs also allow comparison of detected peptide fragments against organism-specific databases to identify the 'parent' protein within the original sample. Unlike RT-qPCR, this is an involved and time-consuming

process, but detection of these proteins provides a definitive answer as to whether a gene is present and translated within an organism's genome.

A mix of these alternative approaches (RT-qPCR and proteomics) is a powerful technique in confirmation of observations made from RNA-sequencing datasets. The *Wolbachia* endosymbiont of insects and filarial nematodes have been the subject of multiple such experiments in recent years. By applying a combination of transcriptomics, RT-qPCR, and/or proteomics, these experiments have focused on attempting to establish, and further understand, the biological basis that underpins the relationship the *Wolbachia* endosymbiont may have with its hosts.

One example of work that utilised qPCR to validate observations within the transcriptome was performed by Strübing *et al.* in 2010[296]. In this study, microarrays were used to investigate the transcriptional responses of the rat filarial nematode *Litomosoides sigmodontis* after treatment with tetracycline for up to 36 days (and consequent depletion of their *Wolbachia* endosymbiont) with follow-up studies utilising RT-qPCR. As specific microarrays were not available for this nematode species owing to a lack of genome sequence, the study utilised microarrays designed for the related human filarial nematode *Brugia malayi*. The approach of Strübing *et al.* highlighted the mitochondrial respiratory chain as being upregulated upon *Wolbachia* depletion by both microarray and RT-qPCR, and whilst the fold-change results of both approaches did not match, the directionality of all test genes did. Specifically, Strübing *et al.* identified the upregulation of co-factors of the respiratory chain, such as cytochrome c oxidase and cytochrome *bc*1[296], after *Wolbachia* depletion. Both of these co-factors require haem to function, a metabolite that has been the focus of multiple studies investigating *Wolbachia*'s relationship with its nematode hosts[111,167,247].

A second example was work done by Bennuru *et al.* in 2016[253], looking at the transcriptome and proteome of the filarial parasite *Onchocerca volvulus* and its *Wolbachia* endosymbiont (*w*Ov) across several life cycle stages. This included adult males, adult females, microfilariae, embryos, and vector life cycle stages. This study validated 75% of *O. volvulus*'s annotated genes as being actively transcribed during at least one life cycle stage, with 64% of these genes also validated as being actively translated. In addition, the work also validated 465 of the 785 *w*Ov proteins that were predicted in its genome[297]. Bennuru *et al.*[253] did not report a perfect match between transcriptomes and proteomes (r value <0.39), but did report good correlation in terms of directionality across the studied life cycle stages (p value <0.0001).

These comparative analyses between transcriptomic approaches with a secondary technique only strengthens findings of biological interest within these *Wolbachia*-nematode host relationship studies. As such, the aim of this study is to apply such secondary approaches to the gene finding and transcriptomics experiments outlined in **Chapter 2 and Chapter 3**. These studies have focused on analysis of the filarial nematode *Brugia malayi*, and its *Wolbachia* endosymbiont (*w*Bm) during the critical L3 to L4 developmental stages within the final mammalian host. The results of these studies include successful identification of 21 new protein-coding genes in *w*Bm that were missed during the original sequencing project[111], 5 instances of non-model transcriptional events that may be occurring within the genome, and an additional 104 genes in *w*Bm that have no identifiable homologues with known function. Additionally, several genes and pathways that showed differential expression in *w*Bm across 4 time-points in *B. malayi* development (Days 3, 7, 11, 14 post-infection) were also identified. Secondary approaches to be used for this work includes RT-qPCR of several newly identified targets and genes to confirm expression, as well as confirm observed differentially expressed transcripts. Proteomics work will also be applied in an attempt to validate the presence of any of the 21 newly identified protein-coding genes.

4.2 Results

4.2.1 Proteomics validation of new annotation

All Proteomics datasets used for analysis was generated by Dr Gemma Molyneux

The updated genome annotation of *w*Bm (as detailed in **Chapter 2**) identified a total of 826 protein-coding genes, made up of 805 genes identified by Foster *et al.* in 2005[111], as well as 21 newly identified protein-coding genes, and 3 functional-RNA genes. To identify if any of the 21 newly identified genes from the reannotation study are actively translated into a functional protein, proteomics analysis was also performed in an attempt to identify *w*Bm translated proteins in *B. malayi* excretory-secretory products.

Publicly available raw mass-spectrometry data from parasitic nematodes are scarce, with the only examples originating from *Onchocerca volvulus*[253], *Onchocerca ochengi*[298], and *Litomosoides sigmodontis*[299]. As a result of this, datasets from a previous study by Molyneux *et al.* (unpublished) were used for analysis. The objective of this previous work was to study the excretory-secretory products of *B. malayi* microfilariae that had been treated with tetracycline to deplete *w*Bm, as compared to untreated controls. Briefly, microfilariae were collected from the intraperitoneal cavities of the Mongolian jird animal model and left for 24 hours in RPMI media to collect excretory-secretory products. After incubation, collected

RPMI media was then concentrated down, before proteins were precipitated, pelleted by centrifugation, and washed with ice-cold acetone. After trypsin digestion and separation by liquid-phase chromatography, these purified peptides were then used as input for Mass Spectrometry (MS) analysis to identify *B. malayi* excretory-secretory products, and any that may show differential expression between experimental conditions. See methods for full details.

As this work intended to look for the presence of translated *w*Bm proteins, the raw results files from MS of untreated *B. malayi* excretory-secretory products were parsed through Sequest[294], using the wrapper program Proteome Discoverer (ThermoFisher, see methods for full search parameters used). This search utilized a custom-made database that concatenated the 21 new protein-coding genes identified by the reannotation study (**Chapter 2**), as well as the 805 genes identified in the original 2005 study by Foster *et al.*[111]. The gene models of these 805 genes had also been refined based on RNA-sequence data (**Chapter 2**). Additional searches against a decoy database were also done to reduce the chances of a random match. These searches utilized the post-translational modifications of Carbamidomethylation, Acetylation, and Phosphorylation.

Gene ID	Protein product function	Protein size (amino acids)	Peptide coverage	Peptide count
wBmNew0004	Uncharacterised membrane-bound protein	88	40.91%	12
wBm0150	Cell division protease FtsH	609	44.99%	31
wBm0188	YidC/Oxa1 membrane protein insertase	572	14.69%	10
wBm0209	Pyruvate phosphate dikinase	889	27.00%	26
wBm0221	Zinc-dependent peptidase	421	44.89%	18
wBm0528	Transcription termination factor Rho	471	42.04%	23
wBm0581	Uncharacterised secreted protein	606	23.10%	15
wBm0732	Integrase/recombinase XerD	328	52.74%	18

Table 4-1: List of 'Master' proteins identified from peptides processed by Sequest. *Note the presence of* wBmNew0004 *within this list, a newly identified gene that encodes for an uncharacterised membrane-bound protein. All listed protein products have at least 1 unique peptide mapping to them from the input spectrum data.*

The analysis identified a total of 9,921 peptides which mapped to a total of 808 out of the 826 *w*Bm proteins used for the input database. From these proteins, 8 'master' proteins were identified that had a combination of the highest peptide coverage, number of unique/cross-mapping peptides to the gene, and confidence of the peptide mapping (**Table 4-1**). All of

these master proteins identified had a minimum of 10 mapping peptides, at least 1 of which is unique (i.e. mapping only to the associated protein out of all *w*Bm proteins used in the input database), with this peptide also passing stringent confidence filter score (ion filter score $p < 0.01$) using a decoy database. Of interest for this study is the presence of peptides mapping to the *wBmNew0004* gene, which strengthens the results made from the reannotation study in **Chapter 2**. This gene had no identifiable function but was predicted to have a transmembrane domain. A local alignment tool, DIALIGN[300] was used to identify the location for where these peptides originate from within the protein, with all of the peptides appearing to originate from the last 40 amino acid residues of the protein (**Figure 4-1**).

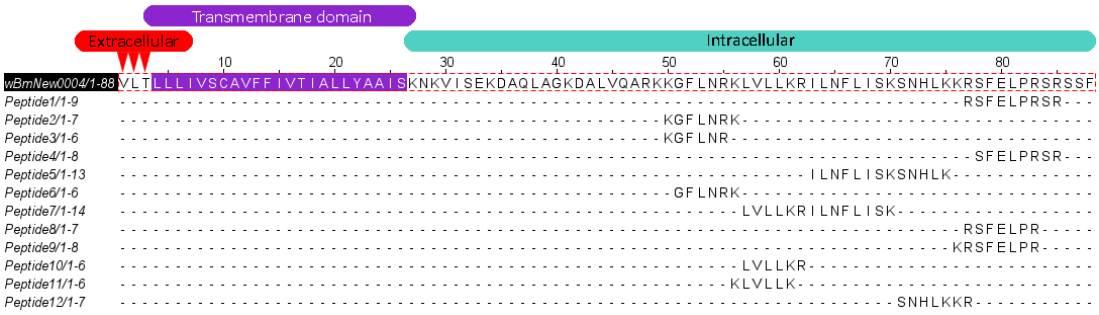


***Figure 4-1: Alignment of the* wBmNew0004 *sequence (highlighted with a red border) with the 12 peptides identified by MS in the secretome of* B. malayi.** *Predicted protein topology annotations are also included at the top of the figure.*

4.2.2 Bioinformatic characterisation of a new translated *w*Bm protein

The identification of a new, translated protein from the reannotation project is of significant interest and could possibly play a role in *w*Bm biology, despite the lack of annotation for this gene (*wBmNew0004*). As such, bioinformatic characterisation of this protein of interest was then attempted. Initial searches indicated the presence of a transmembrane domain, with confirmatory searches using InterProScan[122] confirming the presence of a transmembrane domain between amino acid residues 4 to 26, with residues 26 to 88 predicted to be localised intracellularly. No other protein domains were detectable. BLAST[121] searches with BLOSUM45 of the corresponding protein sequence against the NCBI nr database indicated homologues exclusively within the *Wolbachia* genus, with typically 50% query coverage and between 40-60% sequence identity. These homologues in other *Wolbachia* typically aligned to the first 55% of *wBmNew0004* (up to residue 50), which includes the transmembrane domain, whilst the remainder had no alignment.

A more in-depth comparison to the genome of *Wolbachia* of *Wuchereria bancrofti* (*w*Wb, genome assembly[137] ID ASM220423v2) was also performed using BlastN and manual alignment using T-Coffee[202]. This comparison was done because the *w*Wb genome was noted

to have the highest number of homologues to *w*Bm based on results from the Large-Scale Blast-Score Ratio (**Chapter 2**). This analysis identified a potential homologue with sequence identity of 57%, the highest of the 30 genomes used as input for analysis. Manual alignments of wBmNew0004 to the genome of *w*Wb was able to identify a partial homologue that was interrupted by 2 stop codons, as well as 2 frame-shift mutations at the beginning of scaffold NJBR02000098. Whilst this homologue in *w*Wb could be subjected to non-model translational events, the lack of RNA-sequence data for *w*Wb precludes further analysis. This indicates that wBmNew0004 appears to be unique to *w*Bm.

Subsequent PSI-BLAST[200,201] searches against the remaining 49 residues of the protein with no identifiable domains identified possible, distantly related homologues with kinesin-like, or microtubule-binding domain containing proteins. These hits were noted to have a recurring motif that was conserved within the last 50 amino acids of *wBmNew0004* (**Figure 4-2**). In addition, these hits were noted to be from a wide range of eukaryotic taxa, examples including Mammalia, Viridiplantae, Arthropoda, Nematoda, and Fungi. The individual hits were noted to typically have 77% sequence coverage with 31-36% sequence identity when compared to *wBmNew0004*. The majority of protein hits were over 1,000 amino acid residues long, with a minimum of approximately 200 residues. Interestingly, the first 300 amino acids of these potentially distantly related homologues appear to be well-conserved, with the remainder of the protein being highly variable.

InterProScan analyses of a selection of these PSI-BLAST hits identified the presence of kinesin, and kinesin-motor domains in all of the selected hits, as well as predicted key residues for ATP (8 residues) and microtubule (3 residues) binding. These microtubule binding residues were predicted to be part of the *Eg5* spindle pole protein kinesin motor, a member of the Kinesin-5 family of proteins[301,302] (conserved domain entry ID **CD01364**). Proteins containing this motor domain are known to play a role in spindle formation and chromosome segregation during cell division, during both meiosis or mitosis[301,303,304].

A selection of these PSI-BLAST hits were downloaded and compared against *wBmNew0004* using the local aligner DIALIGN[300], which was chosen for alignment due to the disparity in sequence length between *wBmNew0004* and detected hits. This identified a conserved region within all these hits which aligned to *wBmNew0004*, with this region noted to have a kinesin-heavy domain that also included the 3 predicted microtubule-binding sites (**Figure 4-2**). Within *wBmNew0004*, 2 of these 3 microtubule-binding sites were noted to be conserved, with the final binding site mutated from a Lysine to a Glutamine residue.

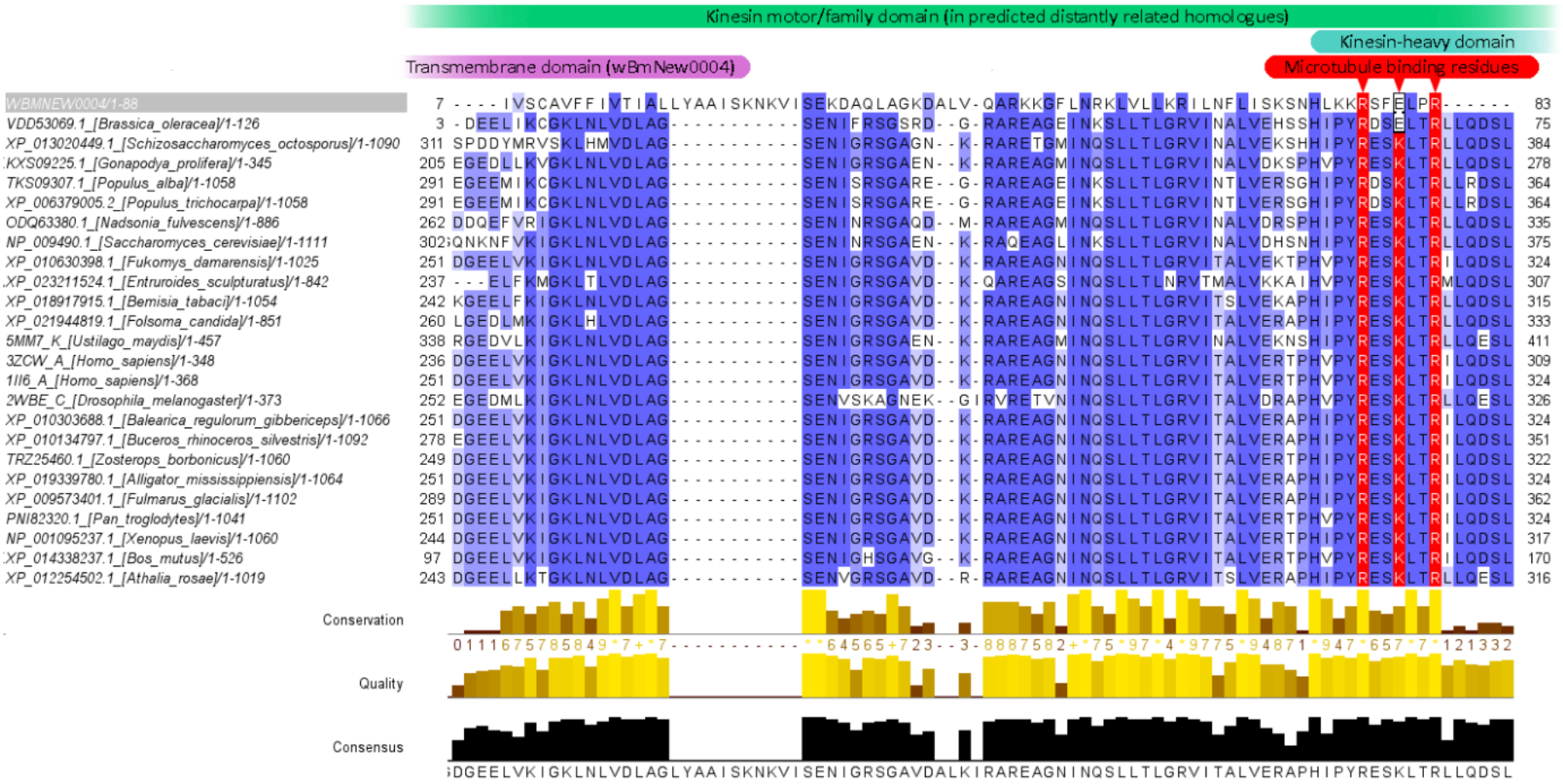


Figure 4-2: Alignment between the translated protein* wBmNew0004 *and kinesin-domain containing proteins from a selection of a wide range of taxa. *Species origin of protein is indicated on the left in square brackets, preceded by protein ID number from either Genbank or the Protein Data Bank. Note the highlighted microtubule binding residues in red which are highly conserved between the selected homologues and the gene of interest, with only one exception. Also note how in the gene of interest there are a number of additional residues in the overall kinesin motor/family domain that are highly conserved across all genes.*

4.2.3 RT-PCR validation of new and existing genes with no known function

Of the 21 newly-identified protein-coding genes, 8 were noted to have no homologues with identifiable function after BLAST searches against existing databases (**Table 2-1**), and only one was detectable via proteomics analysis (*wBmNew0004*). By comparison, within the 805 genes that were originally identified, 104 genes were also noted to have no homologues with known function (**A 1-3**). This makes a total of 112 predicted protein-coding genes in *w*Bm currently with no known function, and no evidence of existence outside of the presence of an open reading frame. In an attempt to confirm expression, and thus presence, for several of these genes, RT-PCR was performed on a shortlist of these 116 genes.

Selection from the 116 genes with no known function was based on two factors: **(i)** their lack of homologues with known function, **(ii)** a general trend of increasing expression level over the 4 studied time-points as determined by FPKM measurements. From this shortlist, 15 genes from the original 2005 annotation, and 2 genes from the reannotation study were randomly selected for amplification, and primers designed. An additional 3 genes of known function in *w*Bm were also included to act as a positive control for expression. These 3 genes of known function were selected based on: **(i)** a general trend of increasing expression level over the 4 studied time-points as determined by FPKM measurements, **(ii)** their low expression level in *w*Bm, using a 5% threshold of FPKMs for all *w*Bm genes (FPKM > 5.4, **Chapter 2**). This list of 20 genes and their function is summarised in **Table 4-2**.

The Day 14 time-point was chosen for amplification due to its high load of *w*Bm, as determined by *wsp*/*GST* ratio experiments (**Chapter 2**), as well as the read-mapping efficiencies seen within the RNA-sequence datasets (**Figure 3-2**). Additionally, prior published experiments[15] also corroborates this time-point as having the highest load of *w*Bm out of the 4 studied time-points. Reverse-Transcriptase (RT)-negative controls were also included to ensure bacterial gDNA was not contributing to amplicon signal.

DNAse treatment of RNA samples collected from the Day 14 time-point, and subsequent reverse-transcription and PCR identified all 20 genes as successfully amplifying (**Figure 4-3**). Three targets continued to show non-specific amplification within the RT-negative samples (samples 10, *wBm0213* sample 12, *wBm0481*, and sample 17, *wBm 767*, **Figure 4-3**) even in DNAse-treated samples, albeit at lower intensity, when compared to the RT-positive samples. Importantly, the two genes identified during the reannotation study amplified successfully from cDNA, with no bands identified in the no RT-control samples.

Gene ID	Product function (if known)	Classification
wBm0066	Predicted/hypothetical gene	Gene of interest
wBm0648	50S ribosomal protein L7/L12	Reference gene
wBm0085	Predicted/hypothetical gene	Gene of interest
wBm0253	Predicted/hypothetical gene	Gene of interest
wBm0576	Predicted/hypothetical gene	Gene of interest
wBm0034	Predicted/hypothetical gene	Gene of interest
wBm0237	NADH Ubiquinone oxidoreductase	Reference gene
wBm0613	Predicted/hypothetical gene	Gene of interest
wBm0414	Predicted/hypothetical gene	Gene of interest
wBm0213	Predicted/hypothetical gene	Gene of interest
wBmNew0017	Hypothetical, membrane-bound protein	Gene of interest
wBm0481	Haem biosynthesis protein HemY	Reference gene
wBm0089	Predicted/hypothetical gene	Gene of interest
wBm0487	Predicted/hypothetical gene	Gene of interest
wBm0587	Predicted/hypothetical gene	Gene of interest
wBmNew0001	Predicted gene of new annotation	Gene of interest
wBm0767	Predicted/hypothetical gene	Gene of interest
wBm0703	Predicted/hypothetical gene	Gene of interest
wBm0193	Predicted/hypothetical gene	Gene of interest
wBm0771	Predicted/hypothetical gene	Gene of interest

Table 4-2: Overview of the 20 genes targeted for amplification. This includes 15 predicted genes with no known function. *Genes included in this table come from from the Foster et al. 2005 annotation of the* wBm *genome*[111]*, 2 predicted genes from the reannotated* wBm *genome with no known function, and 3 genes from the Foster et al. 2005 annotation with known function. ID numbers as listed on the leftmost column correspond to the gel lanes as illustrated in* ***Figure 4-3****.*

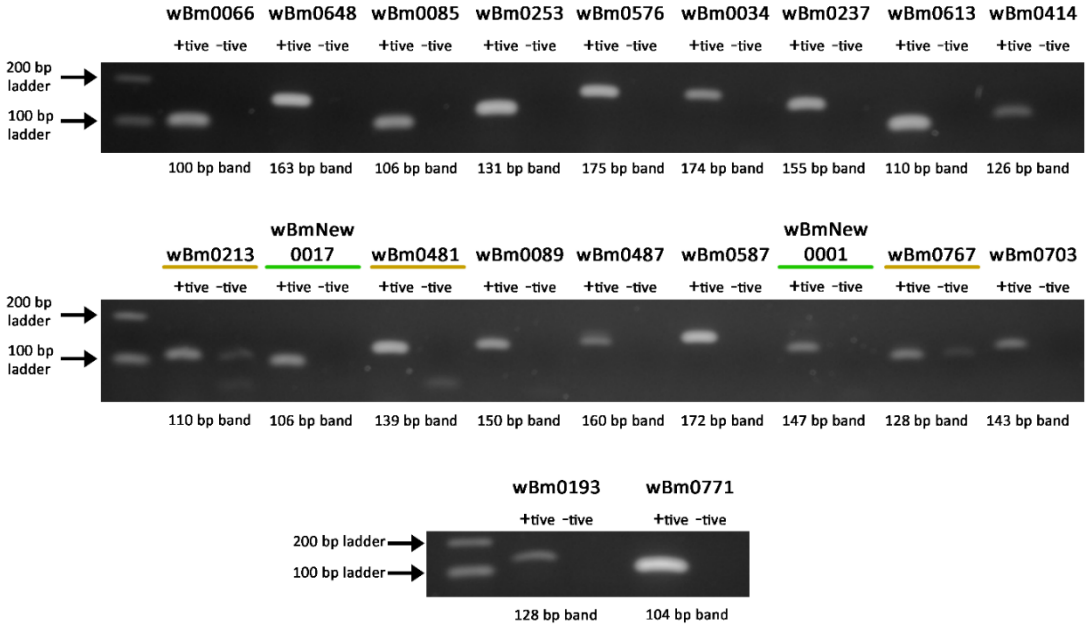


Figure 4-3: Agarose gel electrophoresis of the 20 targets of interest, labelled with Gene ID's as per Table 4-2. *The orange underline under the Gene ID labels indicate the primers that generated non-specific amplification within the reverse-transcriptase negative control, even after DNAse treatment. The green underline under the Gene ID labels indicate the amplicons from the 2 newly identified from the reannotation study in* ***Chapter 2****.*

4.2.4 RT-qPCR validation of non-model translational events

In addition to the 21 newly identified protein-coding genes of *w*Bm, there were an additional 5 cases of non-model translational events, which includes 2 cases of Stop-Codon Readthrough (SCR) and 3 cases of Programmed Ribosomal Frame-shifting (PRF). As these genes were not previously identified as being expressed, RT-qPCR was performed to confirm that they are actively transcribed by *w*Bm. This was done by designing 3 sets of primers (**Table 4-3**) that targeted: **(i)** the 5 genes of interest, **(ii)** 4 intronic regions of the *w*Bm genome that were predicted to have no expression, **(iii)** 4 genes with known function that were predicted to be actively transcribed by observation of RNA-sequence read alignments. The latter 4 genes that were chosen to act as 'positive controls' were identified based on a known function, as well as their low expression level in *w*Bm, using 5% threshold of FPKMs for all *w*Bm genes as established previously (FPKM > 5.4, **Chapter 2**). One of the genes used to act as a control (*wBm0284*) had also been validated as expressed as proteins in previous work, both in *w*Bm and a transgenic yeast model[227]. Quantification cycles (Cq) for all amplicons were measured, and statistically significant differences in means between the 3 groups of Cqs were then calculated using a one-way ANOVA to see if there was any significant differences between any of the 3 test groups. This was followed by a Tukey-Kramer multiple pairwise comparison test to identify which group shows a statistically significant difference in means Significant differences between groups was defined by an adjusted p-value less than 0.05, as determined by the Tukey-Kramer test for pairwise comparison.

Gene ID (if present)	Product function (if known)	Target
wBmNew0101	ATP-dependent Clp protease, ClpA	Gene of interest (SCR)
wBmNew0102	GspD secretin/Type II secretion system	Gene of interest (SCR)
wBmNew0103	Alpha-beta hydrolase	Gene of interest (PRF)
wBmNew0104	ABC transporter protein	Gene of interest (PRF)
wBmNew0105	PQQ repeat family dehydrogenase	Gene of interest (PRF)
Intron (409063 – 409214)	N/A	Negative control
Intron (550605 – 550723)	N/A	Negative control
Intron (671300 – 671479)	N/A	Negative control
Intron (779855 – 780014)	N/A	Negative control
wBm0266	Preprotein translocase subunit SecA	Positive control
wBm0476	tRNA processing exoribonuclease BN	Positive control
wBm0762	Predicted Permease protein	Positive control
wBm0284	*Wolbachia* outer surface protein Wsp	Positive control

Table 4-3: Overview of amplicons identified to confirm the expression of the 5 genes that showed possible instances of non-model translational events. *Amplicons are grouped into 3 different categories:* ***Genes of interest***

(the 5 predicted cases of non-model translational events), ***negative control*** *(an intronic region predicted to not be transcribed), and* ***positive control*** *(a gene with known function predicted to be transcribed).*

All 13 targets of interest showed amplification after RT-qPCR, with the 4 negative control genes showing statistically significant differences when compared to the 5 genes of interest and the 4 positive control genes (**Figure 4-4**). The lack of statistical significance between the genes of interest group and the positive control group, indicates that they have similar transcriptional activity, and that the gene of interest group can be predicted to be expressed.

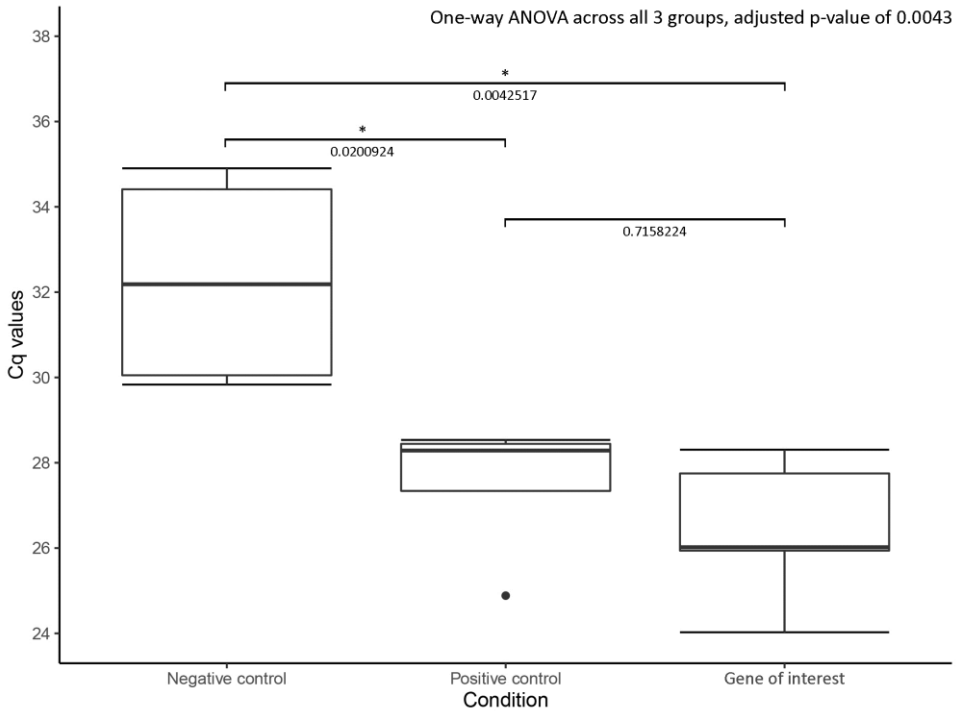


Figure 4-4: Boxplots to show the distribution of Cq values between Gene of interest amplicons, Positive control amplicons, and Negative control amplicons. *P values are annotated between groups tested, with the presence of asterisks indicating statistical significance. Note how the negative control group is statistically significantly different from both the Positive control and Gene of interest groups, whilst the Positive control and Non-model transcript groups show no statistical difference.*

4.2.5 RT-qPCR of differential expression analysis

Differential expression analysis and manual inspection of *w*Bm's transcriptome during the first two weeks post-infection identified several pathways of interest that may contribute to the symbiotic relationship, or at least *w*Bm's population growth. These include the glycolysis/gluconeogenesis pathway, the Tricarboxylic Acid Cycle (TCA), and the glutaminolysis pathway (**Chapter 3**). To confirm these observations, RT-qPCR was performed across the 4 studied time-points (3, 7, 11 and 14 days post-infection) using 13 genes that were randomly selected from the aforementioned 3 pathways (**Table 4-4**). These 13 test

genes were accompanied by an additional 3 reference genes (**Table 4-4**), chosen based on their product (DNA mismatch repair protein MutL, *wBm0396*, large ribosomal protein subunit L20, *wBm0657*, and the membrane protease subunit HflK, *wBm0768*). These 3 reference genes had an EdgeR-calculated fold-change value at or close to 0, and are involved in core processes that would not be expected to show differential expression under these conditions. This comparative analysis method between the 13 genes of interest and the 3 reference genes require that all PCR products are amplifying at the same rate with each thermocycle (known as qPCR reaction efficiency[293,305,306]), or that the efficiencies of each reaction are normalised to each other mathematically[307]. For this analysis of 13 genes, the Pfaffl method was used to mathematically normalise the efficiencies between the 13 genes.

Gene ID	Product function	Associated pathway	Correlation
wBm0207	Pyruvate dehydrogenase, E1-α subunit	Glycolysis	Partial
wBm0209	Pyruvate phosphate dikinase	Glycolysis	Negative
wBm0298	Glyceraldehyde 3-phosphate dehydrogenase	Glycolysis	Negative
wBm0403	Phosphoglyceromutase	Glycolysis	Positive
wBm0666	Pyruvate dehydrogenase, E1-β subunit	Glycolysis	Partial
wBm0244	Malate dehydrogenase	TCA Cycle	Partial
wBm0367	Isocitrate dehydrogenase	TCA Cycle	Negative
wBm0519	Succinyl CoA synthetase, α subunit	TCA Cycle	Negative
wBm0540	Aconitate hydratase	TCA Cycle	Partial
wBm0051	NADPH-dependen glutamate synthase	Glutaminolysis	Positive
wBm0182	Glutamate dehydrogenase	Glutaminolysis	Positive
wBm0443	GMP synthase	Glutaminolysis	Partial
wBm0512	Carbamoyl-phosphate synthetase	Glutaminolysis	Negative
wBm0396	DNA mismatch repair protein, MutL subunit	Reference gene	N/A
wBm0657	Large ribosomal protein subunit L20	Reference gene	N/A
wBm0768	Membrane protease, HflK subunit	Reference gene	N/A

Table 4-4: Overview of genes chosen to be amplified for RT-qPCR and validation of differential expression results from the transcriptome of* wBm *across the first two weeks of* Brugia malayi *development. *A 'correlation' column was included to indicate the correlation in directionality between RNA-seq and RT-qPCR, with results expanded on in* ***Figure 4-5*** *and* ***Figure 4-6****. The colour scheme used for this column is also kept throughout these figures, with green representing full, or positive correlation, blue representing partial, and red/orange representing negative between the two aforementioned techniques.*

Across the 13 genes of interest, a total of 8 genes showed either full (positive), or partial, correlation between the results from RNA-sequencing and RT-qPCR in terms of directionality of fold-change, but not necessarily magnitude. By contrast, a total of 5 genes showed no correlation in directionality between the results from RNA-sequencing and RT-qPCR (**Figure 4-5**, **Table 4-4**). The distribution of genes that do not show any correlation also appears to be randomly distributed across the 3 different pathways studied.

Interestingly, *wBm0051* (encoding NADPH-dependent glutamate synthase) showed good directional correlation between the RNA-sequencing and RT-qPCR data, as did *wBm0182* (encoding Glutamate dehydrogenase), although not necessarily good correlation in magnitude. This indicates that the formation of glutamine from glutamate can be confirmed to be upregulated in *w*Bm after the first week of infection of *B. malayi* (**Figure 4-5**, **Figure 4-6**). In addition, several of the genes that either showed partial, or no, directional correlation between RNA-sequencing and RT-qPCR fold-changes, had RT-qPCR fold-changes that are comparatively very small. Examples of these include *wBm0207*, *wBm0209*, *wBm0298*, *wBm0666* (genes within the glycolysis pathway), and *wBm540* (genes within the TCA cycle) (**Figure 4-5**, **Figure 4-6**).

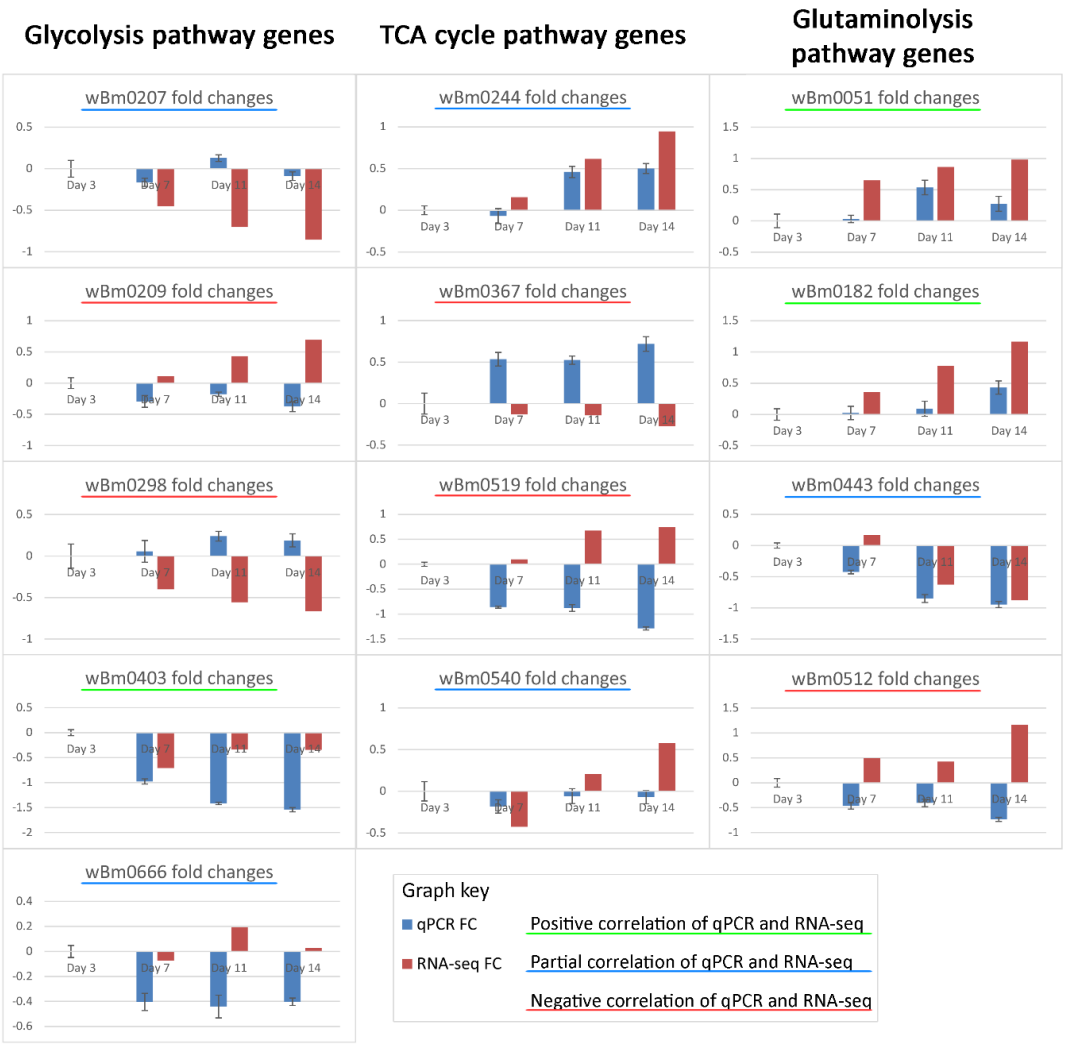


Figure 4-5: Comparison of Log_2 fold-changes of 13 genes of interest calculated from RT-qPCR (blue columns in all samples) versus fold-changes calculated from RNA-sequencing differential expression (orange columns), using Day 3 as a baseline. *Colour of the title underline indicates concordance of direction (gene is up- or down-regulated) between the RT-qPCR fold-changes and RNA-sequencing fold-changes. A green underline indicates good concordance, blue indicates partial, and red indicates no concordance.*

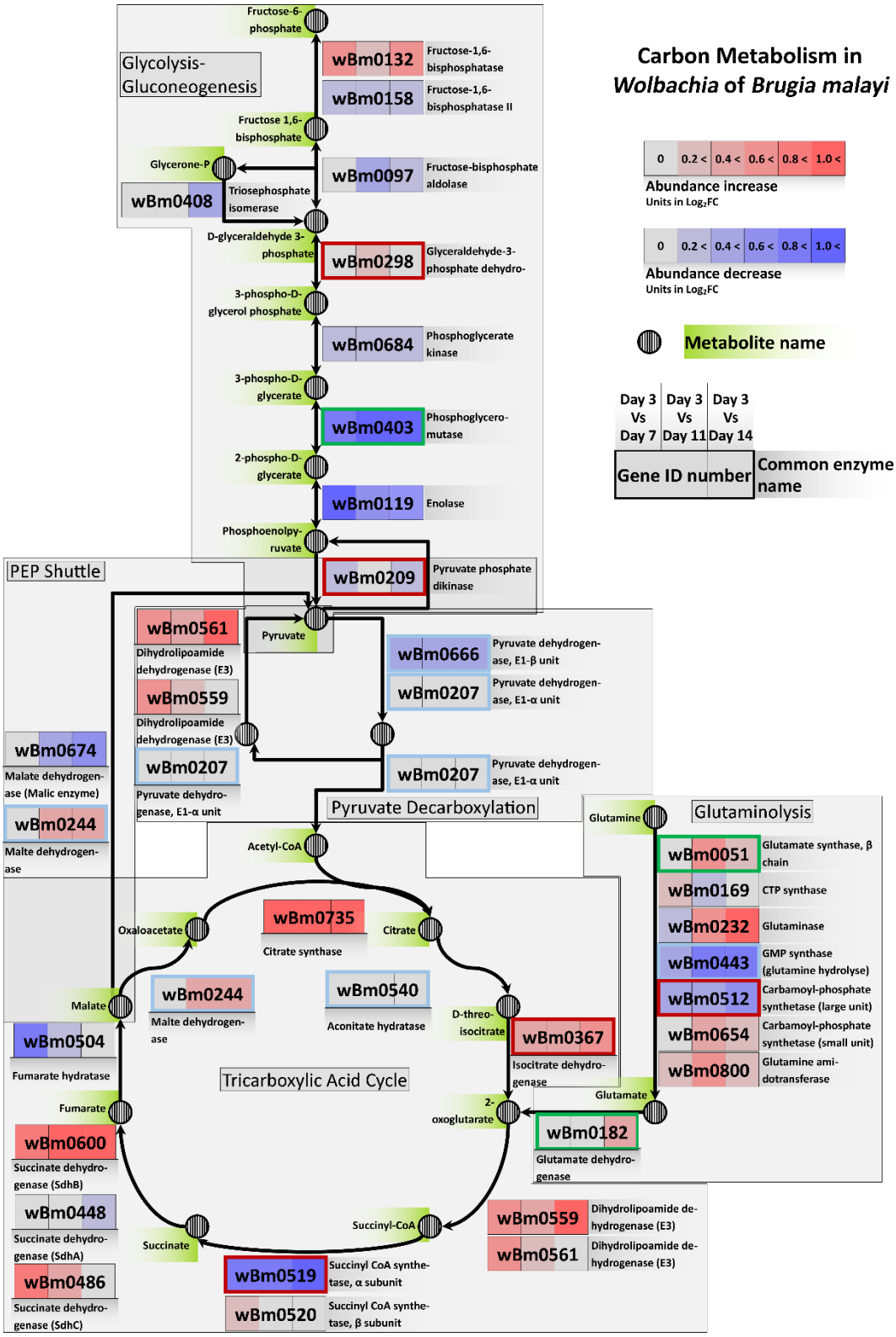


Figure 4-6: The Glycolysis, TCA, and glutaminolysis pathways of wBm, with fold-change results from RNA-sequencing, as well as RT-qPCR, overlaid. *Genes with a bolded outline were sampled for RT-qPCR analysis and have had the new fold-changes overlaid, with the border colour depending on the correlation of results RNA-sequencing and RT-qPCR, reflecting the same colour code used for the title underline in* ***Figure 4-5****. Despite the disparity between RNA-sequencing and RT-qPCR fold-changes seen in* ***Figure 4-5****, the glycolysis pathway is still downregulated, whilst the TCA cycle is upregulated. Glutamate synthase is still considered upregulated, implying that glutamine is actively used by* wBm *just after the L3 to L4 moult of the nematode.*

4.3 Discussion

RNA-sequencing is a powerful tool for the study of gene expression in a variety of conditions, and can lead to the identification of many candidate genes for further study, in addition to biological pathways of interest that require validation by secondary approaches. The application of both RT-PCR-based techniques, as well as proteomics techniques, in this study was successful in confirming the active expression and translation of a number of newly-identified genes within this study. As the proteomics experiments from which the raw data originates from was designed to look at the secretome of *B. malayi* nematodes, the identification of 8 *w*Bm proteins by unique peptides is a significant finding. Although only one of these is predicted to be secreted due to the presence of a signal peptide, the presence of a newly annotated gene during the course of this study (*wBmNew0004*) is a significant finding. Bioinformatic characterisation has identified a potential relationship with kinesin-motor domain proteins, specifically the presence of two residues that could potentially interact with microtubules. Both microtubules and kinesin-motor domain proteins are structures that are only known to be present in eukaryotes[308], with an exception in the *Prosthecobacter* genus of bacteria[309,310].

Kinesin-motor proteins, their structure, and parts of their amino acid sequence, appear to be highly conserved across all eukaryotes, and can be divided into up to 14 different families, possibly more[302,311,312]. These proteins perform a variety of functions within eukaryotic cells, including intracellular transport as well as having roles in mitosis and meiosis within the cell cycle[303]. Despite the large number of families, kinesins maintain a well-conserved structure between all their forms (as reviewed by Marx *et al.*[302]). To summarise, kinesin-binding proteins operate as multimers, usually dimers, with one terminus of the multimer connected to its cargo, followed by a long coiled-coil 'body' connected to a globular 'head' via a 'neck' of coiled α-helices. This head is predicted to contain a conserved binding site for ATP and microtubule, known as the Walker A and Switch-2 motifs respectively[302,313,314]. Although there has been some debate as to the actual 'walking' mechanism that kinesins utilise for migration[315], a general consensus is the use of ATP hydrolysis to achieve motion[316]. Upon binding and hydrolysing ATP in one of the dimer subunits, the globular head is predicted to be rotated forward via the neck linker, before binding to microtubules in front of it and releasing ADP and inorganic phosphate. This is repeated for the second half of the dimer, then the first half again, thus creating a walking motion that propels the protein and its cargo forward along the microtubule[316]. In turn, microtubules are long chains comprised primarily of two protein subunits- α and β tubulin. These subunits bind to each other via GTP (which is

not hydrolysed), and the dimers self-assemble into long microtubule chains via hydrolysis of GTP[317]. Upon construction, the chains remain intact as long as a GTP 'cap' is present at the microtubule end- removal of which results in the deconstruction of the entire microtubule[317].

The structure of kinesin motor proteins in eukaryotes differs significantly from *wBmNew0004*, which maintains very little of the aforementioned conserved structure, such as the ATP binding sites or the coiled-coils of the 'neck' and 'body'. Instead, it appears to be anchored into a membrane, and maintains two residues of the microtubule binding site motif, which appears to be highly conserved in eukaryotes, and may allow *w*Bm to interact with host microtubules in some manner. As *Wolbachia* in general are known to be located within a vacuole derived from the host lipid membranes[29,30], it may be possible that *wBmNew0004* is localised within this vacuole, with the 'tail' that contains the microtubule binding domains effectively localised intracellularly in the host cell. Ever since the first *Wolbachia* genomes were sequenced in 2004 and 2005, one consistent observation was the apparent lack of any flagellar or fimbrial genes that enable motility of *Wolbachia* through, or even in-between, host cells[111]. It was instead believed that *Wolbachia* may migrate through the host cell via co-opting host actin polymerisation, due to the presence of WASP family proteins in *w*Bm which have homologues in the distantly related, obligate intracellular *Rickettsiae*[111]. These WASP family proteins in *Rickettsiae* are predicted to activate the host *Arp2/3* complex, thus stimulating the formation of actin comet tails which have been shown to be utilised by *Rickettsiae* for motility[153,154]. The existence of *wBmNew0004* may thus indicate the use of microtubules by *w*Bm as an additional means of migration through host cells. The similarity of this new gene to the *Eg5* family of kinesins, and their roles in mitotic spindle formation for chromosome segregation, implicates a possible role in spreading through host tissues as cells proliferate during nematode growth. Further experiments will need to be done to formally characterise this, due to the lack of additional protein domains in *wBmNew0004* that contribute to function in the Eg5 family of kinesins. Additionally, questions arise as to why this protein, and its microtubule-binding domain, appears to be unique to *w*Bm and is not shared amongst other *Wolbachia*, despite its possible role in allowing bacterial motility and spread throughout host cells. There may be the possibility of sequencing errors occurring within the genome of *w*Wb interrupting this gene, particularly as it is located next to a scaffold junction. Additionally, this gene may not be properly annotated in other *Wolbachia*, and that manual curation of other *Wolbachia* genomes may be necessary to identify associated homologues.

Outside of this proteomics work, this study was able to confirm the expression of two newly-annotated membrane-bound proteins (*wBmNew0001* and *wBmNew0017*) by RT-PCR. The confirmed expression of *wBmNew0001* is of interest primarily due to its combination of protein domains that are conserved across the majority of *Wolbachia* supergroups (its transmembrane domains), as well as the protein domains that appear conserved only within the individual supergroups (as described in **Chapter 2**). Its localisation within the transmembrane region also raises the possibility of this gene playing a role in interacting with the *Wolbachia* host. This, along with identification of the VirB2 components of the Type IV secretion system proteins in *w*Bm, increases the repertoire of methods that *w*Bm has in interacting with its nematode host. In addition, all 5 incidences of non-model translational events were confirmed as expressed (*wBmNew0101* to *wBmNew0105*) at a level equivalent to *w*Bm's genes that have been previously confirmed to be present and translated into proteins. Both 'test' and positive control groups of amplicons showed statistically significant difference in quantification cycle point compared to the 'negative' control group of amplicons that targeted intronic regions, confirming expression of these genes that exhibit non-model translational events.

RT-qPCR analysis of 13 genes of interest in *w*Bm during the first two weeks of infection (3, 7, 11 and 14 days post-infection) identified 8 genes that showed either a positive (3 genes) or partial (5 genes) directional correlation with results from RNA-sequencing as described in **Chapter 3**. An additional 5 genes had a negative directional correlation between the differential expression analysis and RT-qPCR. This group of 10 genes that either show only partial, or negative correlation, between RT-qPCR and RNA-sequencing data appear to be distributed randomly across the 3 pathways surveyed, indicating no bias in results for a particular pathway in *w*Bm's glucose-pyruvate-glutamine metabolism that could explain the negative correlation in 5 of the genes of interest.

Of the 3 genes that were noted to show positive directional correlation between RNA-sequencing and RT-qPCR, *wBm0051* (encoding NADPH-dependent glutamate synthase) and *wBm0182* (encoding Glutamate dehydrogenase) are involved in the conversion of glutamine to glutamate, and finally to 2-oxoglutarate to be used in the TCA cycle. This validates the overarching theory put forward in **Chapter 3**, in that *w*Bm appears to become more reliant on glutamine for entry into the TCA cycle as time passes. Further to this, despite the 2 glycolysis-related genes that show no concordance to RNA-sequencing fold-change results, the overall trend still indicates a persistent downregulation of the pathway as time progresses during the first two weeks of *B. malayi* infection in a mammalian host, as

illustrated in **Figure 4-6**. This is due to a combination of genes that, despite showing a negative correlation between RT-qPCR and RNA-sequencing results, also showed lower magnitude fold-changes than initially predicted (*wBm0207*, *wBm0209*, and *wBm0298*, encoding pyruvate dehydrogenase β, pyruvate phosphate dikinase, and glyceraldehyde 3-phosphate dehydrogenase respectively). Additionally, genes such as *wBm0403* and *wBm0666* (encoding phosphoglyceromutase and pyruvate dehydrogenase α) showed significantly higher magnitudes in downregulation fold-change from RT-qPCR than initially predicted. Interestingly, *wBm0367* (encoding isocitrate dehydrogenase) of *w*Bm's TCA cycle was noted to be upregulated by RT-qPCR during later stages of infection by its *B. malayi* host. Despite this being the opposite of what was initially predicted by RNA-sequencing in **Chapter 3**, this enzyme has been cited as performing one of the rate-limiting steps within the TCA cycle, and can be actively controlled by phosphorylation in *E. coli* to utilise the glyoxylate bypass[318]. This bypass directly converts isocitrate to succinate and glyoxylate, with the latter able to be further metabolised into malate. While such a bypass does not exist in *w*Bm, upregulation of *wBm0367* during later stages, as well as the upregulation of *wBm0244* (encoding malate dehydrogenase), and the lack of significant up/down regulation with *wBm0540* (encoding aconitate hydratase), indicates that the TCA cycle as a whole is unlikely to be downregulated, despite the significant downregulation of the succinyl CoA synthetase α subunit. Thus to summarize, the observations from RT-qPCR are still broadly in line with the hypothesis put forward in **Chapter 3**. Although *w*Bm may not upregulate its TCA cycle as highly as initially predicted, *w*Bm still indicates a preference for the usage of glutaminolysis for energy generation, and a lack of reliance on the glycolysis/gluconeogenesis pathway.

The application of secondary techniques has strengthened the observations made from the reannotation of *w*Bm's genome. The confirmation of a newly annotated gene, in the form of a translated protein, is of significant interest, particularly due to the location of two microtubule-binding residues within its amino acid sequence, and the possible implications for *w*Bm biology. The identification of newly annotated transcripts, either from RT-PCR or proteomics, yields avenues for further study of the *w*Bm-host symbiotic relationship, or potentially basic *Wolbachia* biology in general due to the presence of non-model translational events.

4.4 Methods

4.4.1 Proteomics sample preparation

1 million *B. malayi* microfilariae were collected from the peritoneal cavity of gerbils via RPMI washes and separated from host material via passage through PD-10 columns. Purified

microfilariae were cultured for 24 hours at 28°C in 3ml of RPMI media. Excretory-secretory products were concentrated to 0.5ml using an Amicon Ultra 3000 MCO device (Milipore). Proteins were precipitated with ice-cold trichloroacetic aid (25% w/v final concentration), and pelleted by centrifugation at 13,000x g. The pellet was then washed 3 times with ice-cold acetone and left to air-dry, before being resuspended in ammonium bicarbonate (25 nM, pH 8.0). Proteins were quantified via nanodrop, then treated with 0.1% v/v RapiGest™ (Waters) at 80°C for 10 minutes followed by reduction with dithiothreitol (DTT) in a final concentration of 3mM (60°C for 10 minutes), and alkylation with iodoacetamide (IAA) at a final concentration of 50mM (dark at room temperature for 1hr). Trypsin was then added at an enzyme:substrate ratio of 1:40, and incubated overnight at 37°C. Surfactant was then inactivated by treatment with 0.1% trifluoroacetic acid (TFA, 37°C for 1h), and peptides recovered via centrifugation at 13,000x g.

Peptides were resuspended in 0.1% formic acid (FA), and initially separated via reverse-phase liquid chromatography (RPLC) using a DIONEX UltiMate™ 3000 LC chromatography system. Peptides were injected onto the analytical column (Dionex Acclaim® PepMap RSLC C18, 2 µm, 100 Å, 75 µm i.d. x 15 cm, nanoViper.), maintained at 35°C and at a nanoflow rate of 0.3 µlmin-1. Peptides were separated over a linear 180 min chromatographic gradient (3-60 % buffer B in 140 min), composed of buffer A (2.5 % acetonitrile: 0.1% FA) and buffer B (90% acetonitrile: 0.1 % FA).

Samples were analysed using an LTQ Orbitrap Velos mass spectrometer using Xcalibur software v2.6 (Thermo Scientific, UK). A data-dependent Top20 collision induced dissociation (CID) data acquisition method was used with intact peptides detected in the Orbitrap at a mass resolution of 30,000. Ions were scanned between 350-2000 m\z in positive polarity mode. The ion-trap operated with CID MS/MS (with wide band activation) on the 20 most intense ions. Dynamic exclusion was enabled to avoid repeatedly selecting intense ions for fragmentation and this was set at 500 with an exclusion duration of 20.0 seconds. Charge states of 1 were rejected. The minimum MS signal threshold was set at 500 counts and the MS/MS default charge state was 2 with a 1.2 m/z isolation width, normalised CID at 35V and an activation time of 10 ms.

4.4.2 Protein identification

Protein identification was performed via the Sequest HT[294] search engine, via Thermo Proteome Discoverer (version 1.4.1.14). This identification process used a custom database, which contained the amino acid sequences of newly identified proteins from the reannotation study in **Chapter 2**, as well as all existing proteins of *w*Bm as identified by Foster

et al.[111] in 2005. Within this latter group, 25 protein sequences were revised based on revised gene models from the RNA-sequence work, as described in the reannotation study in **Chapter 2**. Genes that encoded for any identified functional RNAs or tRNAs were excluded from this database, as they are unlikely to form a full protein via translation. Raw spectrum files were imported into the software and utilised the following settings for Sequest HT parameters: Maximum of 2 missed tryptic cleavage sites permitted, a precursor mass tolerance of 0.5 Daltons and a fragment mass tolerance of 0.3 Daltons. Carbamidomethyl (C) was set as the only static modification, with additional dynamic modifications: Phosphorylation (S, T, Y) and Methylation (C, I, K L). A decoy database, generated by reversing the sequences of the starting database, was also utilised with stringent and relaxed peptide confidence filters set to 0.01 and 0.05 respectively.

4.4.3 Bioinformatic characterisation

Potential homologues of gene of interest *wBmNew0004*identified as part of the proteomics survey were identified using NCBI's BlastP[121] search, using the BLOSUM45 algorithm against the non-redundant protein sequences database, limited to the *Wolbachieae* organism taxa (taxa ID 952). Subsequently, BlastN[319] searches using the nucleotide sequence of *wBmNew0004* were performed against the genome of *Wolbachia* of *Wuchereria bancrofti* (*w*Wb) to identify genomic regions that may have homology with *wBmNew0004*. The genome was downloaded from the NCBI's Assembly[137] database (assembly ID ASM220423v2). These genomic regions in *w*Wb were translated into proteins, and manual alignments of the different reading frames were performed against the protein sequence of *wBmNew0004* with T-Coffee, and visualised in JalView[320].

The characterisation of genes of interest that were identified as part of the proteomics study was first performed via analysis by Pfam and InterProScan[122] to identify potential protein domains. The online programs Phobius and TMHMM[186,199] were used to confirm and determine the orientation and localisation of the protein, with PSIBlast[200,201] searches used to identify the presence of conserved residues and structures within the proteins of interest. For results found in PSI-Blast, alignments were made using the local alignment program DIALIGN[300] to formally identify the presence of conserved regions and active sites in proteins of interest. Alignments were then visualised and annotated in JalView[320] to confirm the conservation of residues between proteins of interest and distantly related homologues identified via PSI-Blast.

4.4.4 Primer design

All targets that were used for (q)PCR amplification during this study were identified based on FPKM measurements of RNA-sequence alignments to the *w*Bm genome, quantified using Cufflinks[189,321], or by differential expression analysis by EdgeR[127] and pathway analysis by PathView[258]. Target regions, either genes or intron regions, were analysed using the online tool Primer-BLAST[322], with the following changes to default settings: PCR product size maximum of 200 base-pairs, Primer melting temperatures minimum of 58°C, maximum of 62°C, and maximum melting temperature difference of 2°C. Predicted targets and amplicons were then BLASTed against the nr database, organism filtering for *Wolbachia* endosymbiont strain TRS of *Brugia malayi*, taxid:292805. All primers that were finalised and used for reactions are listed in **Table 4-5** through to **Table 4-7** within their respective sections.

4.4.5 DNAse treatment of RNA samples and cDNA synthesis

B. malayi samples depleted of rRNA via Terminator Exonuclease (Epicentre) that were previously extracted from nematodes 3, 7 (2,000 nematodes each), 11 and 14 days (1,000 nematodes each) were collected and concentration of remaining samples measured on a Nanodrop spectrophotometer. Sample collection of these nematodes was described in **Chapter 2**. From this initial concentration, samples from all 4 time-points were diluted down to a concentration of 20ng/µl. Using this diluted sample, DNAse treatment was performed using BaselineZERO DNAse treatment (Epicentre), following manufacturer's protocol. Briefly, 2µl of DNAse reaction buffer and 1µl of DNAse enzyme was added to 17ul of suspended RNA, with the samples incubated at 37°C for 30 minutes. Subsequent to this, 2µl of 10x DNAse stop solution was added, and the samples incubated at 65°C for 10 minutes.

DNAse treated samples from the 4 time-points of interest were then taken forward for cDNA synthesis using SuperScript Reverse Transcriptase IV (Invitrogen), following manufacturer's protocol. Briefly, 11µl of template RNA was mixed with 1µl of kit-supplied 50ng/µl random hexamers and 10mM dNTP mixes, before centrifugation and heating at 65°C for 5 minutes. To this, 4µl of SuperScript IV buffer, 1µl of 100mM DDT buffer, Ribonuclease inhibitor, and SuperScript IV Reverse Transcriptase was added. This mixture was then incubated at 23°C for 10 minutes, then subsequently 55°C for 40 minutes, and 80°C for 10 minutes. To this mixture, 1µl of *E. coli* RNAse H was then added to remove the presence of remaining mRNAs, and incubated at 37°C for 20 minutes. For all 4 time-points, a negative reverse-transcriptase control (RT-control) was also included, substituting the SuperScript IV Reverse Transcriptase with molecular grade water (Ambion).

Additional cDNA, as well as an additional negative RT control, was also generated from *B. malayi* total RNA, collected from 600,000 microfilariae (mf), with this cDNA used for calibration of primers and confirmation that they work. Briefly, mf in RNA-Later were first diluted via adding an equal of phosphate-buffered saline to the sample, then centrifugation at 2,000x g at 4°C for 10 minutes to pellet mf. The supernatant was removed, before 1ml of TriZol Plus reagent (Invitrogen) was added. These samples were then transferred to separate sterile 2ml screw cap tubes containing ceramic beads of different sizes (1.4 and 2.8mm) (CKMix, Bertin). Batches were homogenized at 6,000rpm (Minilys, Bertin Instruments) for 4 x 30 seconds, cooling on wet ice for 30 seconds in between.

200µl molecular grade chloroform (Invitrogen) was then added for every 1ml of TRIzol reagent, and mixed by vigorously shaking the tube for 15 seconds, followed by incubation at room temperature for 5 minutes. The sample was then centrifuged at 12,000x g for 15 minutes at 4°C. The upper aqueous phase was removed and transferred to a nuclease-free 1.5ml tube. An equal volume of ice cold 70% ethanol was added to each sample and vortexed for 2 - 3 seconds. The tube was inverted several times to disperse any precipitate.

Subsequent binding, washing and elution of RNA from the sample was then conducted using Exiqon's miRCURY™ RNA Isolation Kit – Cell & Plant, in accordance to manufacturer's instructions. gDNA was also removed via an on-column treatment using Baseline-ZERO™ DNase (Epicentre). Final RNA was eluted in 2 x 50µl molecular grade water (Ambion).

4.4.6 RT-PCR of primers checking for expression of predicted/hypothetical genes

Both Reverse Transcriptase (RT)-positive samples and negative-RT controls obtained from the Day 14 time-point were taken for PCR of the 20 gene targets, using Platinum PCR Supermix (Invitrogen). Briefly, 1µl of cDNA generated from the 600,000 mf was mixed with 45µl of Platinum PCR Supermix (Invitrogen), and 2µl of 10mM concentrations of forward and reverse primers for the associated amplicon of interest, generating a 0.4mM concentration of both forward and reverse primers. PCR mixes were then incubated for 95°C for 15 minutes, then 40 cycles of 94°C for 30 seconds, 62°C for 30 seconds, and 72°C for 1 minute, before a final incubation of 72°C for 1 minute. These mixes were then run on a 1% Tris-Borate EDTA agarose gel for 1 hour 30 minutes at 120 volts and 80 miliAmps to allow migration, using a 100 base-pair ladder for measurement. Gels were then imaged and photographed via a UV transilluminator gel documentation system (Syngene) via GeneSys software to confirm presence of bands.

Gene ID	Forward primer	Reverse primer
wBm0066	AGATCAAAGAAGAAACAGGCAGG	CCATACATTGTGACATAAGCACCA
wBm0648	ACCTACTAGTGATAATGCTGCCA	GGATTCGACTAACTCTTTTGCTTCT
wBm0085	CACACCCATCTGGTAAGTCCT	GTTGTGTGGAAATGCCTGAA
wBm0253	GGCAAAATCTTTACACGAAGCG	GAGCAGAGAGAACTTGTTGTGC
wBm0576	GGATGACGATGTTTTGTACCTTGA	ACGTCCTTTTCTTCTCTGTGAGT
wBm0034	TCAGTTTCAGTCCCGAGTAAACT	TGGTTAACACATATATCAAGCAGCG
wBm0237	GCAGCATTCCCAGAAGCAATAA	TCGGTTACACATATCGGCTGAA
wBm0613	AGCAACTTATGGCTGGTGGTAT	GTCAATCATTTCCTTCATCTGAGCA
wBm0414	CGCACACTCTCTTTTCTGCATG	TCATAGGAGTCATGATGAAGGTGT
wBm0213	ACCTACGATCTTCAAGAGCTGC	CCATATTTTCCATGTGTCGGGG
wBmNew0022	AAAGATGTTGGAGCAAAAGCGC	GTTCACAAATAAATGGCGCACA
wBm0481	GGCTGTGAAGTCTTTATTTAGGGC	GTTTACTGTGTACTCCGCCTCT
wBm0089	AGCGGGAACACTCATAGAAACA	CTCCTTTGTCTCTCCCTTCCAC
wBm0487	GCATAGTATACCCCTATTCAAGCCA	CCATATATCCGGAGTAATTCTTGTGC
wBm0587	GCGATATACTGCAAAAGCGTGA	AACTTTGCAGGAACCACAACAC
wBmNew0001	TTGCAGCTCTATTCATTGCAGC	TTACAGCCTGGGTAACTTGGTC
wBm0767	GGACAAGTAGCAAGGGAGGAAA	ACCATTGAATCTTGCAGCCCTA
wBm0703	CAGCGGGTTATCAGGAAGAAGT	TGGGTCCATCACATTATGCCAA
wBm0193	ACCACAATCCGATAGCAAAGGT	TTCCGTACTCCAGCTGTTTTGT
wBm0771	CTTCATGAACAGCCGCTAACAC	GACCGATCACTACCTTCACCTG

Table 4-5: Overview of primers used for validation of gene models as described in Chapter 2. *Please see* ***Table 4-2****, for more information on the particular gene amplicon.*

4.4.7 RT-qPCR of non-model translational events

Selection of intronic regions to act as a negative control was based on visualisations in the Artemis genome viewer[190,191], and subsequent identification of non-coding regions of the *w*Bm genome that had little to no RNA-sequence alignments (**Table 4-6**).

Samples were run on a BioRad Chromo 4 machine using 20µl reaction volumes, each of which contained 10µl of Quantitect SYBR Green mix (Qiagen, UK), 300 nM primers, nuclease-free H2O and 1 µl of cDNA input. Cycling parameters for these reactions used a BioRad C1000 Thermal Cycler, and consisted of: a hold at 95°C for 15 min, 40 cycles of 94°C for 15 sec and 62°C for 30 sec, and extension at 72°C for 1 minute. Fluorescence was monitored during the 72°C step, using the FAM channel. Melt curve analysis was performed between 60°C to 97°C at a ramp rate of 0.2°C/sec. All samples were repeated in triplicate, with one negative control spiked with RT-negative sample included for each. Statistical significance between the three groups was determined in the R programming language via a one-way ANOVA test, followed by Tukey-Kramer test for multiple pairwise comparisons to identify which pair showed statistical significance. The results of this were then plotted in R using the ggplot2 package. Code for significance tests is as follows:

my.res.aov <- aov(Cq.values ~ Condition, data = values)

TukeyHSD(my.res.aov)

Gene ID	Forward primer	Reverse primer
wBmNew0101	AAATGTGCTTAGGTCGAGCGT	CCAGCTCCAACGATAGTGTGT
wBmNew0102	AGTACATCGCACAGCAGAGT	TGACTTCTGACTTAATGCGGCT
wBmNew0103	AACTGTCCTGTGCGTTTATTGC	ATTCCAACGAACGGTTGTCAG
wBmNew0104	GACTGACAGGATCATGGGTCG	GCGTCTGCGAGACGTGTTAT
wBmNew0105	ATGGTCCAAGCCACTAAAGGT	GACGTCCAGACTACTTCTCCG
Intron (409063 – 409214)	TTGCCTTTGGGGTTCCTAGTT	TCACTAAATAGCACCGAACAGAAGA
Intron (550605 – 550723)	AGACAAAGGAAGGTGATTGGGAT	CTCCATCATTGCAGGCTCTCA
Intron (671300 – 671479)	GATAACTGCCGCTTGTCAGC	CCTCTTTCCTCTGTCGTCCC
Intron (779855 – 780014)	ACTATAACCGCAATCAGTGGGAA	TGCTGGAAATAATGCTGTGCG
wBm0266	TTATCCGACGAGGAGCTTGC	GAAACCTGCGAGAGGCTTCA
wBm0476	TCCGTTTCTCATTGTCTTAATGGC	AAACTTTGTGGTGGCCCTGA
wBm0762	ATGCGATGATCCTAGGCTTTCTT	TAGCATCATTCCCAGCGTAGC
wBm0284	GGCCTGCATACCAAGTGAAAG	CCTGGCTTACACGGCTACC

Table 4-6: Overview of primers used for amplification of targets to validate the presence of non-model translational events. *Please see* ***Table 4-3*** *for more information on the particular gene amplicon.*

4.4.8 RT-qPCR validation of differential expression analysis

Efficiencies for the 16 identified targets were analysed using standard curves. These standard curves were obtained by using 'pure' amplicons of the 16 targets, which were obtained by regular PCR using Platinum PCR Supermix (Invitrogen), and subsequent gel extraction using the QIAquick Gel extraction protocol (Qiagen) following manufacturer specification. PCR was performed as described earlier. After running the amplicons on a 1% TBE agarose gel, the bands were excised using a sharp scalpel and a UV box, with the excised bands placed into separate Eppendorf tubes, labelled for each of the 16 amplicons of interest.

The excised bands are weighed (after removing tube weight), and 3 volumes worth of gel-weight of buffer QG added, the volumes unique to the weight of the excised bands. This mix of gel and buffer QG was then incubated at 50°C to allow the gel to dissolve, with inversions of the tubes every 2 minutes. Once the gel was fully dissolved, 1 volume worth of gel-weight of isopropanol was added, and the contents mixed by inversion. All samples were then transferred to separate QIAquick spin columns and centrifuged for 1 minute. 750µl of buffer PE was then added to the columns, before centrifugation for 1 minute twice to ensure removal of residual wash buffer. The QIAquick spin columns were then transferred to a new collection tube, before 50 µl of molecular-grade water (Invitrogen) was added, and the columns left to stand for 5 minutes, and centrifugation performed to elute purified cDNA. All centrifugation steps described were run at 13,500g speeds. The eluted purified cDNA was

then quantified on a Nanodrop 3.0 spectrophotometer (Thermofisher), and DNA copy numbers calculated based on the formula below:

$$\frac{x\ ng\ \times 6.0221 \times 10^{23}}{(N\ \times 660)\ \times\ 10^{9}}$$

Where x is equal to the amount of DNA as measured by Nanodrop spectrophotometer (Thermofisher), and N is equal to the length of the amplicon used per sample. This then results in a quantification of amplicon copies per µl of eluted sample.

Using the estimated DNA copy numbers, samples were then diluted to a concentration of 10^{10} DNA copies per µl to act as a working stock. These working stocks were then diluted to a concentration of 10^{7} DNA copies per µl as a starting concentration, before 6 serial 1:10 dilutions were performed for each of the 16 amplicons of interest. These were then used to create a standard curve for each sample by qPCR in triplicate using 20 µl final volumes, with reaction mixes containing: Quantitect SYBR Green mix (Qiagen, UK), 300 nM primers, nuclease-free H2O and 2 µl of DNA. Cycling parameters for these reactions used a BioRad C1000 Thermal Cycler, and consisted of: a hold at 95°C for 15 min, 40 cycles of 94°C for 15 sec and 60°C for 30 sec, and extension at 72°C for 1 minute. Fluorescence was monitored during the 72°C step, using the FAM channel. Melt curve analysis was performed between 60°C to 97°C at a ramp rate of 0.2°C/sec. Quantification cycles (Cq) for each sample were then used to calculate PCR efficiencies using BioRad's CFX Manager program. If the standard curve for any of the 16 samples produced efficiencies below 60%, new primers were ordered and the process repeated.

Once efficiencies for each primer pair was established, all 16 amplicons of interest were tested in triplicate on the cDNA samples obtained from the 4 separate time-points, as described previously. Fold-changes observed by qPCR for each gene was then calculated via use of the Pfaffl method[307], and compared to the fold-change results as calculated by EdgeR from **Chapter 3**.

Gene ID	Forward primer	Reverse primer
wBm0051	AGAACCCGTGAATGTGCCAA	GGACCGAGCCCAACAACTAA
wBm0182	GCTGACCTATGCGTTCTTGC	TCGAACCTCAGCAGCGATTT
wBm0207	TGAGGAAAAGGCAGGGCAAT	AGCGTAAGTCCCAACTGCAA
wBm0209	GAAGGGACTGCCAGCTTCTC	TGTTCTGCAGCTTTTTCCGC
wBm0244	TTGGAATGGCTGGAGTGCTT	AACGAATCAGCGGCACCATA
wBm0298	ACATGCCTACACGAATGACCA	TTGCTGCTCCAGTTGTGGTT

wBm0367	TAACGCTCGCTTGGGACAAA	ATCGCCAGCCCTTTTGAGTA
wBm0403	AAAAGCTGTGCTGGCTGTTG	AGAGTGTGTGCTGTGTGAGG
wBm0443	GAGGCAAGTGGAGTAGCACA	TGCTGCCCATAGCATATCCC
wBm0512	CTTTGACTCTGCGTGATGCG	AACATTTGCACCACCTGCAC
wBm0519	CCGTTGCTGCGGTTTTATCC	CACAACAACAAGTGGCGGAG
wBm0540	ACTGGGACACGAAAAGCACT	GGGGAAATGTGGTCTGTGGT
wBm0666	TGGGTGAAGAAGTTGCGGAG	TCCAACAGCAAGACCAGCAA
wBm0396	ATCCTCTAGGGTTTGCACGC	GGCTGCATGCTGGTCAACTA
wBm0657	TTAATGCAGCAGCGAGGGAA	ACTCCAGCAAGTGCAAGACC
wBm0768	GCTCGTGCAGATAAAGAGCG	ATTGCCTCACCCTTTGCTCG

Table 4-7: Overview of primers used for amplification of targets as described in this study. *Please see* ***Table 4-4*** *for more information on the particular gene amplicon.*

Chapter 5 The role of *Wolbachia* in transmission of mf into the mosquito vector

5.1.1 Introduction

Past studies that focus on evaluating the symbiotic relationship between filarial nematodes and their *Wolbachia* endosymbiont have mainly focused on later life-cycle stages of the nematode[168,169,249,250,253], specifically during L4 to adult developmental stages. This is primarily due to the deleterious effects that are known to occur to adult and/or developing nematodes upon *Wolbachia* depletion, which includes arrested development at the early L4 stage onwards[80,248], a block in embryogenesis[82], and accelerated death of adult nematodes[4,81], thus effectively terminating propagation. Conversely, comparatively little work has been done to investigate the symbiotic relationship between filarial nematodes and their *Wolbachia* endosymbiont during early larval stages. Specifically, such stages include the L3 to L4 development in the final host after vector transmission (covered in **Chapter 3**), and microfilariae (mf, or L1-stage nematodes) stages of the nematode life cycle, and their development within the intermediate mosquito vector.

The life cycle of the human parasitic filarial nematode *Brugia malayi* and its *Wolbachia* endosymbiont (*w*Bm) is complex. This parasitic filarial nematode requires an intermediate mosquito vector to transmit between hosts, as well as to allow development from the mf stage to the infective L3 developmental stage. These mf are released in large numbers daily, each of which maintain a small *w*Bm population (approx. 200 per nematode, as determined by McGarry *et al.*[15]), and are encased within a chitinous sheath - the remains of the nematode eggshell. Depending on the strain of *B. malayi*, these mf populate the blood stream of an infected host in a periodic, or sub-periodic, manner[323,324] that coincides with the peak feeding time of local mosquito populations. Successful colonisation of a permissible mosquito vector[10–12,325,326] requires the nematode to penetrate the peritrophic matrix (if present) and gut wall of the vector. This is then followed by migration to mosquito flight muscle tissues for further development to infective L3 stages[13,327,328]. Thus, the major physical barrier to vector infection for the nematode is the gut wall and peritrophic matrix, of which the latter is known to contain a high abundance of chitin, the primary scaffold and support material in arthropods[39,329,330]. Experimental models for *B. malayi* frequently use the permissive *Aedes aegypti*, Liverpool black-eyed strain, as an experimental mosquito vector[13,326,331]. In this vector, *B. malayi* is known to successfully penetrate the gut wall before the formation of a peritrophic matrix[332,333], which occurs 4-8 hours after the blood meal. As such, in this vector, it can be assumed that the major physical barrier for *B. malayi* is the mosquito midgut lining

itself, rather than the peritrophic matrix. Shedding of the microfilarial sheath is known to occur during this penetration process, although different studies make contrasting observations as to whether this occurs before or after penetration[10,12]. Throughout this entire process, *w*Bm populations remain static in mf, and continue to remain static until L3 stage nematode infection of the final mammalian host[15].

During this development within the mosquito vector, there has been a long-standing assumption that *Wolbachia* plays a minimal role in nematode development. This has been attributed to the static, low-level population of the *Wolbachia* endosymbiont during these same life cycle stages[15] (**Figure 5-1**). Indeed, removal of *Wolbachia* in mf has been shown to have little to no deleterious effects on their viability in blood[85,279].

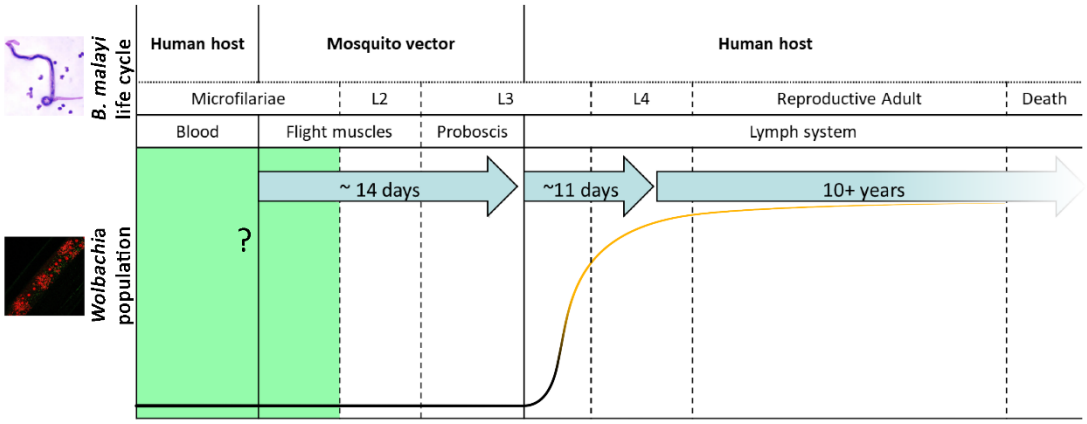


Figure 5-1**: Graphical representation of the life cycle of* **B. malayi, and the concurrent population changes with its* **Wolbachia** ***endosymbiont, not drawn to scale.*** *Developmental life cycle stage of interest for this study is highlighted in green. Note how the* Wolbachia *population stays stable at comparatively very low levels during this developmental stage, and the subsequent stages within the mosquito vector (mf-stage, through to L3 stage). Images of* Wolbachia *taken from Taylor et al.*[67]*, and* B. malayi *taken from the Centers for Disease Control and Prevention*[70]

However, there is a body of work spanning the previous 15 years that serves to challenge the assumption that *Wolbachia* in mf are 'dormant' passengers during these developmental stages. The earliest such work to directly analyse *Wolbachia*'s effects on mf development was done by Arumugam *et al.* in 2007[85], which looked at the rat parasitic nematode *Litomosoides sigmodontis* that utilises the mite *Ornithonyssus bacoti* as its arthropod vector. Using the Mongolian jird animal model (*Meriones unguiculatus*), Arumugam *et al.* found that after 6 weeks of tetracycline treatment and subsequent *Wolbachia* removal, L3 recovery rates from dissected mites 14 days post-feed was significantly reduced in nematodes depleted of *Wolbachia* (9%, vs 54% in control). This is despite equivalent mf concentrations in the gerbil blood, and thus mite inoculation levels, between the treated- and untreated gerbil groups[85]. Reimplantation of these recovered L3s into uninfected gerbils resulted in

skewed adult sex ratios amongst treated nematodes as compared to controls, with more males than females. Treated male nematodes were noted to have a lower *Wolbachia* population than untreated nematodes, although female nematodes from both groups had comparable *Wolbachia* populations. The study however did not investigate whether the reduced L3 recovery rate was due to an inability to develop from mf to L3, or an inability of the nematode to surpass innate barriers within the mite[85].

More recent work has been done by Albers *et al.* in 2012 using the human parasite *Onchocerca volvulus*[86], human volunteers, and natural infection of wild *Simulium damnosum* blackflies. In this study, infected human volunteers were exposed to blackflies for parasite inoculation, with fed flies captured and reared in captivity for 7 days to allow *O. volvulus* development to the L3 stage. A subset of volunteers were then treated for 6 weeks with doxycycline to remove *Wolbachia*, before all volunteers were again exposed to blackfly bites once a month for the next 5 months, again all fed blackflies were captured after each exposure. Using this system, Albers *et al.* were able to show that statistically significant differences were observed in recovered numbers of L3 nematodes beginning from 3 months post-treatment. After this time-point, blackflies that fed on the treated group showed significantly higher numbers of L1- or L2-stage nematodes, and significantly lower L3-stage nematodes, as compared to untreated groups, thus implying *Wolbachia*'s presence is necessary for *O. volvulus* development to the infective L3 stage[86].

Recent studies at the Liverpool School of Tropical Medicine by Cook *et al.* (unpublished work), investigated if a similar phenotype is present in the filarial parasite *B. malayi*. By treating infected gerbil animal models for 2, 4, and 6 weeks with tetracycline, they were able to show that *B. malayi* nematodes depleted of their *Wolbachia* endosymbiont lose the ability to penetrate the gut walls of the mosquito vector *Aedes aegypti*, black-eyed Liverpool strain, thereby blocking transmission at its earliest stage. This phenotype appears to be based on a dose-dependent manner, with mf treated for 2 or 4 weeks showing slight decreases in ability to successfully infect and develop within the mosquito vector. These two treatment groups showed 8 and 39% *Wolbachia* population reduction respectively. But treatment for 6 weeks, and subsequent reduction of *Wolbachia* population by 89%, effectively blocked the ability of mf to penetrate the mosquito midgut (unpublished work).

These experimental observations may have significant ramifications on previously accepted theories on the nematode-*Wolbachia* relationship, and points to an even greater importance of the *Wolbachia* endosymbiont with regards to the nematode's developmental life cycle.

Thus, the objective of this study is to investigate the relationship that exists between *B. malayi* and its *w*Bm endosymbiont during early mf developmental stages just before vector uptake, and identify the mechanisms by which *B. malayi* loses its ability to successfully infect the mosquito vector. To achieve this, we have generated RNA-sequence data from a treated, or untreated, population of *B. malayi* mf, using a peritoneal infection animal model. These treated or untreated populations of mf were further split into two groups: One for blood feeds or direct injection into mosquitoes to observe their development, and a second group for RNA-sequencing and differential expression analysis. This combination of both experimental observations as well as transcriptomic evidence will thus yield insights into this previously understudied life-cycle stage of *w*Bm-nematode symbiosis.

5.2 Results

5.2.1 Animals used, and Tetracycline treatment outcome

All work with animals as presented in this section was done by Mr. Andrew Stevens, Mr. John Archer, and Miss Amy Marriott.

To investigate the effects of *w*Bm depletion on the transcriptome of *B. malayi* during mf stages just before colonisation of the mosquito vector, as well as to repeat experiments to confirm the previously observed phenotype of *B. malayi*'s inability to penetrate the mosquito vector after *w*Bm depletion, we generated biological material using a peritoneal infection of a gerbil animal model (*Meriones unguiculatus*).

A total of 9 gerbils were infected with L3-stage nematodes and left for a minimum of 16 weeks to allow *B. malayi* development into adult nematodes and production of mf. Subsequent to this, 6 gerbils were fed 2.5% tetracycline hydrochloride in drinking water for 6 weeks, with the remaining 3 gerbils acting as untreated controls. Mf were acquired after treatment via necropsy of all 9 gerbils, generating a total of 6 samples from individual gerbils- 3 treated, and 3 untreated (**Table 5-1**).

We observe a marked reduction in the number of mf collected from the treated arm of the study as anticipated due to blockage of embryogenesis. Within the 3 gerbils in the untreated group, we observed approximately 2 million mf per gerbil. By contrast, with the 6 gerbils in the treated group, we observed two gerbils that yielded no mf, and three with reduced mf loads ranging from 250,00-442,080 mf per gerbil (**Table 5-1**).

Untreated		Treated	
Gerbil number	mf recovery	Gerbil number	mf recovery
1	2,377,332	1 and 2	270,000
2	1,999,666		
3	2,000,459	3	250,000
		4	442,080
		5	0
		6	0

***Table 5-1**: **Overview of mf recovery from infected gerbils, separated between Treated- and Untreated- arms of the group.** Note that in the Treated group, the mf from gerbils 1 and 2 had to be pooled due to the low numbers from initial predictive counts to ensure sufficient recovery.*

Microfilariae source	wsp:gst ratio	Microfilariae source	wsp:gst ratio
Untreated Gerbil 1	0.193406	Treated Gerbil 1 & 2	0.008255
Untreated Gerbil 2	0.18623	Treated Gerbil 3	0.00933
Untreated Gerbil 3	0.258194	Treated Gerbil 4	0.009596
Untreated Average	0.21261	Treated Average	0.009061
Effective reduction in wsp:gst ratio			95.7%

***Table 5-2: Overview of qPCR results from recovered mf, separated by originating gerbil**. Results are displayed as a ratio of* wsp *to* gst *copy number, which measures* Wolbachia *population levels relative to nematode cell population.*

Following on from this, we conducted qPCR analysis to confirm the depletion of the *Wolbachia* population in mf. Briefly, DNA was extracted from collected mf and used for qPCR amplification with primers designed to target the *Wolbachia* surface protein (*wsp*) and *B. malayi* Glutathione-S Transferase (*gst*) genes, as described in McGarry et. al.[15]. By normalising *wsp* to *gst*, this method directly quantifies *w*Bm populations relative to nematode DNA, thus allowing comparisons between two samples. Results from qPCR showed a 95.7% average reduction rate of the *Wolbachia* population in the treated samples, as compared to untreated samples (**Table 5-2**).

5.2.2 Assay of L3 development in mosquito vector

All work with mosquitoes as presented in this section was done by Dr Darren Cook, either as part of this study or previous work.

Previous work by Cook *et al.* (unpublished work) have shown that *B. malayi* mf are unable to penetrate the midguts of susceptible *Aedes aegypt i*mosquitoes after 6 weeks of treatment with tetracycline, and a resultant 89% depletion of the *w*Bm population. After 2 or 4 weeks of treatment, a partial block in the ability of mf to penetrate the mosquito midgut can be observed, indicating a potentially dose-dependent relationship (**Figure 5-2**).

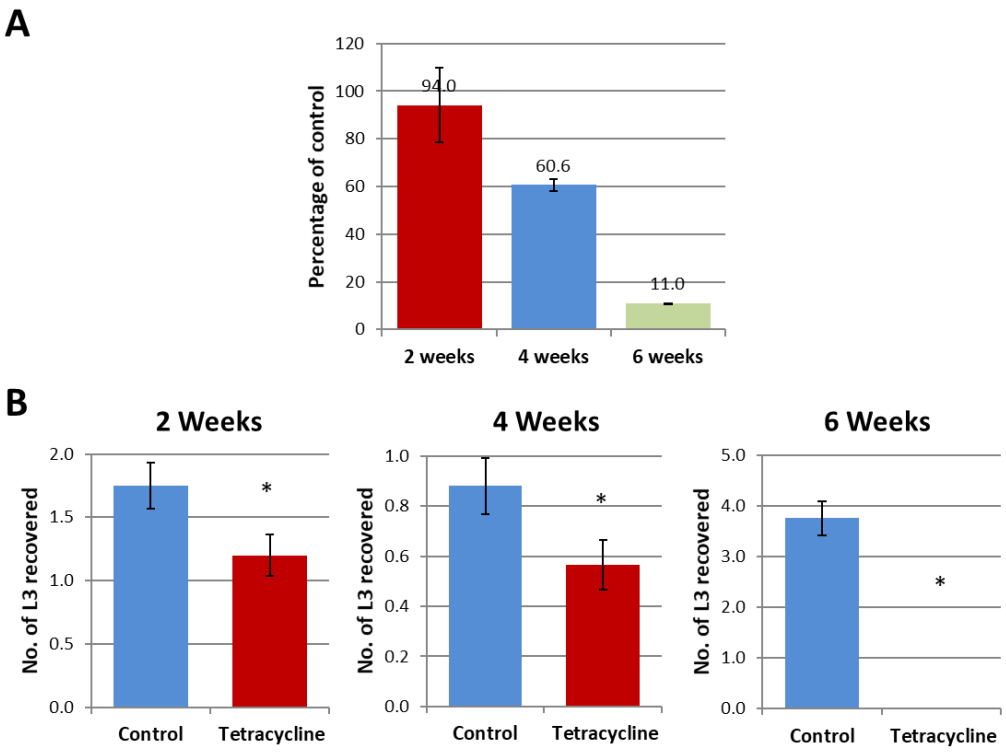


Figure 5-2: (A) Reduction in* wBm *population within mf recovered from gerbils fed 2.5% tetracycline in drinking water for 2, 4, and 6 weeks, measured by wsp:gst ratios. (B) Recovery rates of L3* B. malayi *nematodes from* A. aegypti *(Liverpool black eye strain) mosquitoes by dissection after 2, 4, and 6 weeks of treatment with tetracycline. *Figure taken from Cook* et al. *(unpublished).*

To validate these previous observations and confirm that any differential expression changes within the nematode transcriptome is related to a biological phenotype, treated- and untreated- mf were inoculated into human blood and fed to *Aedes aegypti* mosquitoes (black-eyed Liverpool strain, 20 mosquitoes per group). Mosquitoes were left for 14 days before L3 nematode development was assessed by dissection and microscopy. As a comparison, and to test if the effects of *w*Bm depletion only affected the ability of mf to penetrate the mosquito midgut, we also performed direct injection of treated- and untreated- mf into the mosquito haemocoel (black-eyed Liverpool strain, 16 mosquitoes per group), before subsequent incubation and dissection. Direct injection effectively bypasses the midgut penetration step that occurs during natural infections.

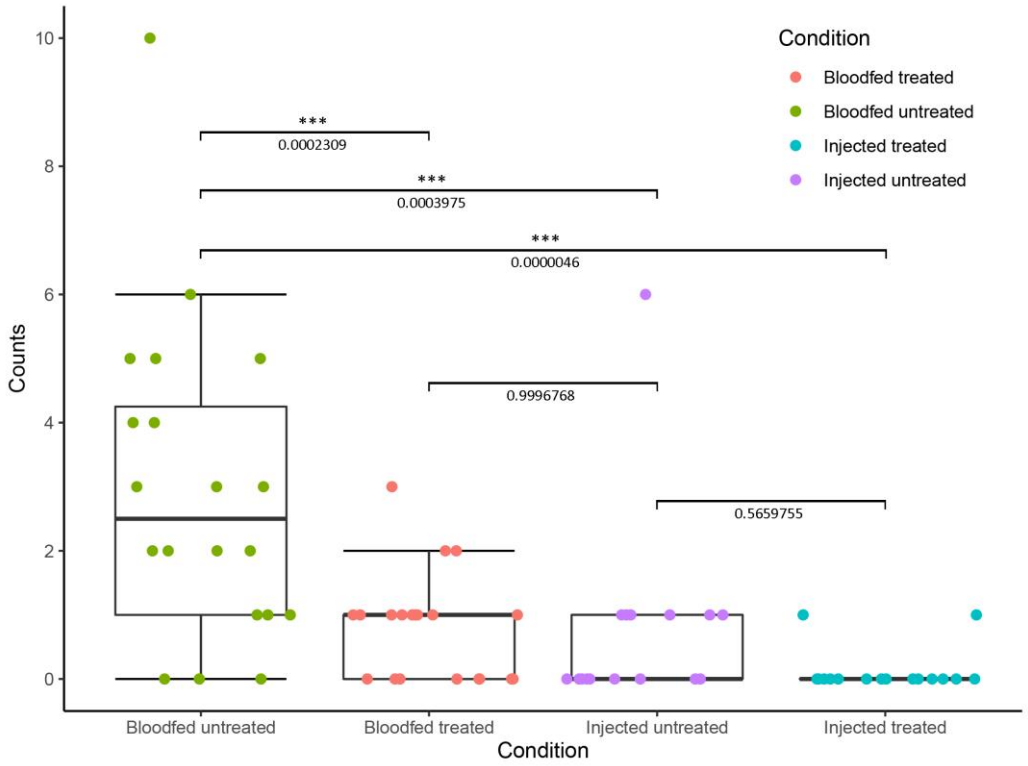


Figure 5-3: Boxplot of the distribution of L3 numbers recovered from mosquitoes according to the 4 experimental arms of mosquito infection, with ANOVA test statistics assessing the impact of treatment overlaid. *20 mosquitoes were used in both blood-fed groups, and 16 mosquitoes used in both injected groups. Note the significantly reduced number of L3s recovered after feeding mosquitoes treated or untreated mf (leftmost 2 columns), and associated statistical significance scores as calculated by a Type I sum of squares two-way ANOVA and Tukey-Kramer multiple pairwise comparisons test. Statistically significant differences would have a value less than 0.05. Also note the low number of L3s recovered after injection of mf directly into the mosquito haemocoel (rightmost 2 columns), and the approaching significance.*

Using a Type-I sum of squares two-way ANOVA, followed by a Tukey-Kramer test for multiple pairwise comparisons to account for different sample sizes, statistical significance was observed in L3 recovery rates between mosquitoes fed treated or untreated nematodes, demonstrating that *w*Bm presence is required for successful midgut penetration and L3 development (**Figure 5-3**, leftmost two columns) in line with previous observations. L3 recovery rates between mosquitoes injected with treated or untreated nematodes was shown to be approaching statistical significance ($P < 0.1$) (**Figure 5-3**, rightmost two columns). This lack of statistical significance may be due to a reduced rate of development of injected untreated mf, as compared to bloodfed controls.

5.2.3 RNA extraction, sequencing, and mapping

Processing and sequencing of RNA samples in this section (with exception of Terminator Exonuclease treatment) was performed by the Liverpool Centre for Genomic Resources

For RNA-sequencing and differential expression, total RNA extraction was performed on microfilariae from the 6 samples of treated and untreated gerbils (**Table 5-2**). Following on

from this, terminator exonuclease treatment was performed on all 6 total RNA extracts. This was done to preserve any *w*Bm transcripts that may be present, whilst removing rRNA species within the sample.

Despite the same procedure and equal amounts of input total RNA, Terminator exonuclease treatment yielded a statistically significant difference in amounts of recovered RNA ($P < 0.05$, two-tailed student's T-Test assuming equal variance). Samples originating from treated nematodes yielded an average of 7% greater amounts of rRNA-depleted RNA as compared to untreated nematodes (**Table 5-3**). The reasons for this discrepancy may be due to the removal of the *w*Bm endosymbiont in the treated group effectively reducing the total amount of rRNA that can be removed from the sample. A second reason may be due to tetracycline's known mode of action of binding to rRNAs and blocking protein biosynthesis, with this binding reducing the efficiency of rRNA depletion, although this should not be possible with eukaryotic rRNAs at concentrations used.

Mf sample origin	Input Total RNA (ng)	RNA recovered (ng)	% after treatment
Untreated Gerbil 1	605	72.8	12%
Untreated Gerbil 2	605	82.2	14%
Untreated Gerbil 3	605	91.6	15%
Treated Gerbil 1 & 2	605	139.0	23%
Treated Gerbil 3	605	111.0	19%
Treated Gerbil 4	605	121.0	20%

Table 5-3: Overview of RNA before and after rRNA depletion. Note how there is a mean difference of 7% between the treated and untreated group of microfilariae, with untreated group having reduced total recovered RNA as compared to the treated group.

All 6 samples were sequenced as paired-ended samples (2x150) using an Illumina HighSeq 4000 at the University of Liverpool's Centre for Genomic Resources. All sequencing samples and read pairs passed quality control (FastQC[245]), with a resultant minimum of 137 million trimmed, unaligned reads per replicate. We then adopted the same strategy as described in **Chapter 3** to separate *B. malayi* reads from *w*Bm reads for accurate differential expression analysis. With each replicate, a minimum of 95.31% of the total reads mapped to the *B. malayi* genome. The remaining unmapped reads were then extracted and remapped to the *wBm* genome. The remaining reads that mapped to the *wBm* genome within the treated group averaged 0.18%, whereas the untreated group averaged 13.07%.

5.2.4 Overview of differential expression analysis

Pairwise differential expression analysis was conducted on the *B. malayi* transcriptome, between the treated and untreated samples. This was done using the EdgeR package[127] being

applied to the raw read counts obtained from FeatureCounts[252]. From this, we identified a total of 1,079 genes (out of a total of 11,535, **Figure 5-4**) that show statistically significant differential expression (FDR < 0.05), with 501 showing upregulation, and 578 showing downregulation. The magnitude of fold-change ranges from -4.8 to 6.4 on a $\log_2$ scale. BLAST searches and cross-checks with existing databases, such as the Kyoto Encyclopaedia of Genes and Genomes (KEGG), were then conducted in an attempt to ascribe function to these genes. The full set of differential expression results is available in **Appendix 4**.

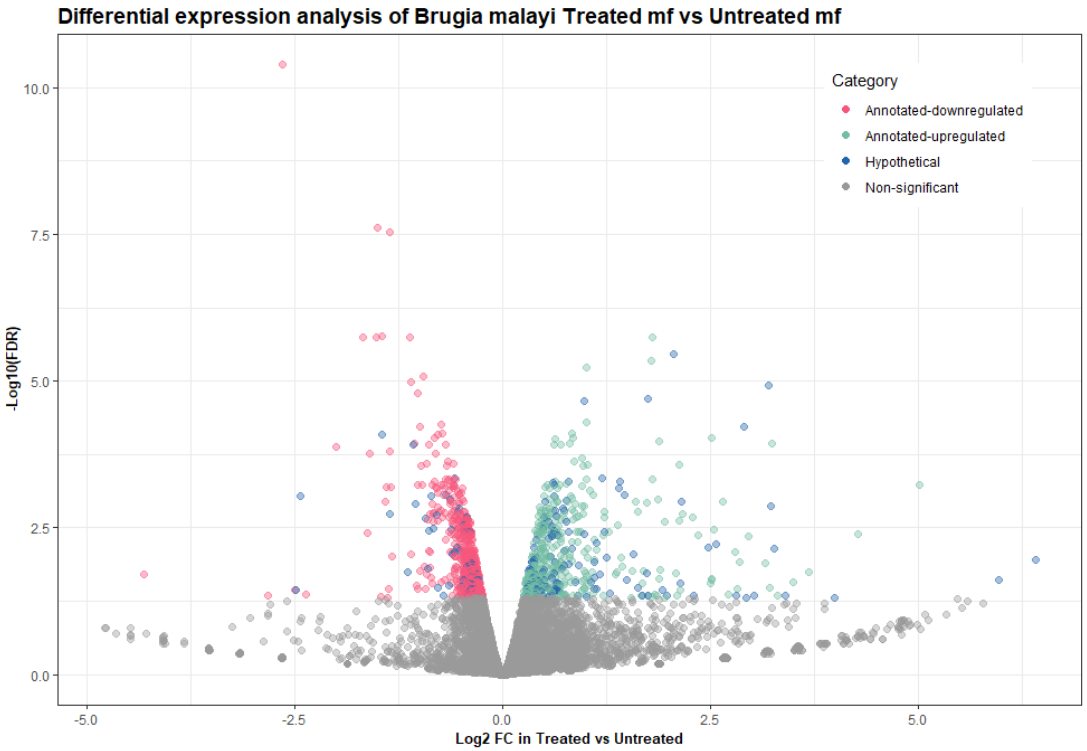


Figure 5-4: Volcano plot of the 1,079 differentially expressed genes in treated mf as compared to untreated mf. *Genes that do not show statistical significance (FDR < 0.05) are coloured in grey. Genes that show upregulation are located positive of the x-axis (green or blue), with genes that show downregulation located negative of the x-axis (red or blue). Note the higher abundance of blue points within the upregulated set, as well as their frequently high $\log_2$ fold-changes.*

Within the 501 genes identified as statistically significantly upregulated, it was noted that nearly 25% of these genes (total 124, **Figure 5-4**, blue points) were identified as having no identifiable function, and/or no identifiable functional domains. Some of the most highly upregulated genes that had a predicted function included an excretory/secretory precursor protein Juv-p120 (*Bm1_29760*), an RNAse H and integrase-like protein (*Bm1_24275*), a Krox-like protein (*Bm1_14700*), and 2 gag-like proteins (*Bm1_21715*, *Bm1_24265*).

The identified excretory/secretory protein Juv-p120 appears to be unique to the *Onchocercidae* genus by PSI-BLAST searches, and has been noted to be secreted by juvenile

female *Litomosoides sigmondontis* nematodes (a filarial parasite of rodents), with this protein showing cross-reactivity for human antibodies that target the microfilarial sheath[334,335]. Whilst no known function has been ascribed to this protein, studies have indicated possibly immunomodulatory effects that may help microfilaria evade host immune systems[336]. Separately, while these genes are not well-characterised in databases, PSI-BLAST searches of the identified gag-like proteins and RNAse H and integrase-like proteins did identify potential homologues in retrotransposons or retroviruses that have been integrated into the genomes of various other mammalian parasitic nematodes. Examples of these parasitic nematodes include *Dracunculus medinensis* (the guinea worm), *Ancylostoma ceylanicum* (a species of hookworm), and *Toxocara canis* (the dog roundworm). These potential homologues frequently had 75% to 100% query coverage, as well as between 35% to 60% identity. Their apparent integration into the *B. malayi* genome, presence in multiple other nematodes, and their lack of homology to genes of prophage WO, a bacteriophage of interest in studying *Wolbachia* and their phylogeny, indicates that these retrotransposons/retroviruses are nematode-specific. Finally, the Krox-like protein is noted to have an *LpxD* domain that is commonly found in UDP-3-O-acyl-glucosamine N-acyltransferase. This enzyme is a component of Lipid A biosynthesis, normally found in plants or gram-negative bacteria. Upregulation of such transcripts may be a response to *w*Bm's depletion, in an attempt to help sustain the population.

Within the 578 genes identified as statistically significantly downregulated, a total of 64 genes had no identifiable function, and/or no identifiable domains. Genes with an identifiable function that showed the greatest downregulation include an Acyl CoA binding protein (*Bm1_21135*), a Never In Mitosis gene A (NIMA)-related Kinase 8 (*Bm1_55750*), a phosphopantothenoylcysteine decarboxylase (part of the coenzyme A biosynthesis pathway, *Bm1_12615*), an excretory/secretory precursor protein Juv-p120 (*Bm1_18010*), and a cuticle collagen precursor (*Bm1_50895*).

The presence of Juv-p 120 excretory/secretory precursor proteins in both the upregulated and downregulated sets is of interest, with alignments of the two genes indicating a 96% query coverage of the two genes, as well as 55% identity. The two identified Juv-p 120 excretory/secretory proteins may elicit different immunomodulatory responses, with the observed changes in expression in microfilariae possibly related to preparation for successful colonisation of the vector host[336]. Outside of these, the NIMA-related kinase plays a role in cell cycle regulation, whilst the cuticle collagen precursor plays a role in nematode moults.

The presence of these two genes within the downregulated gene set may be related to *B. malayi*'s ability to develop within the mosquito vector.

5.2.5 Gene Ontology terms are enriched only in downregulated gene set

For the purpose of Gene Ontology (GO) enrichment analysis, all 1,079 genes that were identified as significantly differentially expressed by EdgeR were split into upregulated or downregulated gene sets based on a $\log_2$ fold-change of > 0.1 or < -0.1 respectively- the minimum $\log_2$ value that differentially expressed genes were identified. Blast2GO[182] was used for the analysis of enriched GO terms, with an FDR cut-off of < 0.05 established as per default program settings. Following this, GO terms were reduced to their most specific level via Blast2GO[182] **Table 5-4**. The full, un-summarised results can be found in **Appendix 4**.

No GO terms were identified as enriched within the upregulated gene set of 501 genes. The lack of enriched GO terms may be explained by the relatively large abundance of hypothetical genes within the upregulated gene set, or possibly indicates that the remaining 377 genes with an ascribed function were poorly characterised. The lack of upregulated GO terms does however indicate that the mf used in this study did not exhibit any stress responses as a result of *w*Bm depletion, mirroring past observation with respect to their viability[85,279]. In comparison to the upregulated gene set, the 578 genes that were identified in the downregulated gene set yielded a total of 110 enriched terms. The number of GO terms was reduced to 23 upon reduction to the most specific terms by Blast2GO (**Table 5-4**). The full list of 110 enriched terms is included in **Appendix 4**.

Three of the most statistically significant terms related to ATP hydrolysis-coupled proton transport (GO:0015991), proton-transporting ATPase activity (GO:0046961), and proton-transporting V-type ATPase, V_1 domain (GO:0033180) (**Table 5-4**). A fourth related term, proton transporting V-type ATPase, V_0 domain (GO:0033179) was also enriched (**Table 5-4**). The V-type ATPase complex that is being referenced is frequently known to utilise ATP to pump protons across membranes and against proton gradients, primarily for the acidification of vacuoles, which is used in a variety of biological processes. The most prominent examples of function include acidification of intracellular compartments, which have a role in generating suitable pH levels for enzymes to function optimally[337], either as part of the lysozyme or possibly generation of conditions for protein maturation. Other examples of function include uptake of cations via H^+-driven antiporter function, or control of synaptic vesicles[337].

Two additional GO terms that showed statistically significant downregulation were associated with nematode larval development (GO:0002119), and oogenesis (GO:0048477) (**Table 5-4**). The former term relates to any processes that allow progression of the nematode larval life cycle over time, from first formation (mf) to eventual mature structure (adult)[131,132,185]. The latter term relates to any process that results in the maturation of ovums or female gametes[131,132,185]. The presence of these two terms indicates that *w*Bm may play a direct role in nematode development, as compared to just impacting the ability for the nematode to penetrate the vector gut wall.

Finally, GO term enrichment also identified two terms that were related to chitinase activity (GO:0004568), and chitin catabolic processes (GO:0006032) (**Table 5-4**). Literature evidence suggests that many filarial nematodes maintain chitinase enzymes that are critical for nematode development, and are expressed at specific life cycle stages that vary based on nematode[338–340]. *B. malayi*'s stage-specific chitinase is known to be expressed only in mf stages of the life cycle[341], and has been hypothesised to be used by the nematode to successfully penetrate the gut wall of the mosquito[341].

Go Category	GO ID	GO Name	FDR Value
Biological Process	GO:0015991	ATP hydrolysis coupled proton transport	4.91E-09
Molecular Function	GO:0046961	Proton-transporting ATPase activity, rotational mechanism	2.13E-05
Cellular Component	GO:0033180	Proton-transporting V-type ATPase, V_1 domain	0.001503
Molecular Function	GO:0030515	snoRNA binding	0.003391
Biological Process	GO:0006364	rRNA processing	0.004901
Biological Process	GO:0002119	Nematode larval development	0.00778
Cellular Component	GO:0005865	Striated muscle thin filament	0.007857
Biological Process	GO:0048477	Oogenesis	0.014148
Cellular Component	GO:0005884	Actin filament	0.020691
Biological Process	GO:0019915	Lipid storage	0.024343
Biological Process	GO:0006754	ATP biosynthetic process	0.024674
Cellular Component	GO:0005774	Vacuolar membrane	0.030055
Biological Process	GO:0006032	Chitin catabolic process	0.031241
Molecular Function	GO:0004568	Chitinase activity	0.031241
Cellular Component	GO:0033179	Proton-transporting V-type ATPase, V_0 domain	0.031241
Biological Process	GO:0006898	Receptor-mediated endocytosis	0.0344
Cellular Component	GO:0033290	Eukaryotic 48S preinitiation complex	0.0344
Cellular Component	GO:0016282	Eukaryotic 43S preinitiation complex	0.0344
Biological Process	GO:0040039	Inductive cell migration	0.035157
Biological Process	GO:0007111	Meiosis II cytokinesis	0.036655
Molecular Function	GO:0008135	Translation factor activity, RNA binding	0.036655
Biological Process	GO:0006633	Fatty acid biosynthetic process	0.041062
Biological Process	GO:0001731	Formation of translation preinitiation complex	0.042605

Table 5-4: List of GO terms enriched within the downregulated gene set of* B. malayi*'s mf transcriptome just before uptake into mosquitoes. *Note the 5 terms relating to Proton-transporting ATPase activity, and the V-type ATPase in particular, and how 3 of these terms are highly significant. Also note the presence of 2 terms related to both chitinase activity and nematode larval development/oogenesis in this list.*

5.2.6 Biological pathways are enriched only in downregulated gene set

To gain a better understanding of the biological relevance that significantly differentially expressed genes may have to biological pathways in *B. malayi*, pathway enrichment analysis, using a similar strategy as outlined in **Chapter 3** with the Generally Applicable Gene Enrichment program (GAGE[135]), was applied. To summarise, all genes that showed statistical significance, regardless of up- or down-regulation, from EdgeR analysis were used as a 'test' set, with their $\log_2$ fold-change data overlaid onto a 'background' gene set, defined as biological pathways by the Kyoto Encyclopaedia of Genes and Genomes (KEGG[255]). This overlay is then used to determine if a given pathway shows an enrichment of up- or down-regulated genes as compared to background, using an FDR of < 0.1, which is a standard threshold used for similar programs. Similar tools such as the Database for Annotation, Visualisation, and Integrated Discovery (DAVID) or Gene Set Enrichment Analysis (GSEA)[134,342] were designed primarily for analysing biological pathways in humans or diseases such as cancer, using data obtained from microarray experiments. GAGE utilises existing *B. malayi*-specific databases, and has been designed to be compatible with RNA-sequencing experimental designs[135].

Using this method, GAGE identified 5 pathways enriched with downregulated genes. These pathways include the Oxidative phosphorylation pathway (*bmy00190*), Phagosome pathway (*bmy04145*), and Lysosome pathway (*bmy04142*) (**Table 5-5**). Additionally, the mTOR pathway (*bmy04150*) showed an FDR score of 0.10013 for enrichment with downregulated genes, close to our statistical significance threshold. In agreement with the output of GO term enrichment analysis, no pathways were identified as enriched within the upregulated gene set (**Table 5-5**).

Downregulated pathways	
Pathway name (ID)	FDR
Oxidative phosphorylation (*bmy00190*)	1.16E-04
Phagosome (*bmy04145*)	4.99E-04
Metabolic pathways (*bmy01100*)	1.14E-03
Protein processing in the endoplasmic reticulum (*bmy04141*)	5.64E-03
Lysosome (*bmy04142*)	3.54E-02
mTOR signalling pathway (*bmy04150*)	1.00E-01

Table 5-5: List of pathways that were identified as enriched with downregulated genes as compared to background. *Note the inclusion of the mTOR signalling pathway, which was very close to meeting the significance threshold.*

5.2.7 Chitinase-activity genes are downregulated

The enrichment of GO terms related to chitinase is potentially significant, as *B. malayi* mf are known to express chitinase enzymes during mf stages of development only, and in no other developmental stages. These chitinase enzymes have been previously hypothesised to aid *B. malayi* infection of the intermediary vector[341], thus their downregulation after *w*Bm treatment is of significant interest.

This study identified 5 genes with potential chitinase function encoded within the *B. malayi* genome (*Bm1_06755*, *Bm1_17035*, *Bm1_28610*, *Bm1_28620*, *Bm1_33145*), of which the 4 underlined were noted as being statistically significantly downregulated in samples that have had their *Wolbachia* population depleted (minimum $\log_2$ fold-change of -0.63, FDR < 0.05). Although the $\log_2$ fold-change of these genes may not be high, the downregulation of 4 out of the 5 chitinase genes of *B. malayi* will likely have cumulative impacts on the nematode.

5.2.8 The V-Type ATPase complex is downregulated

To further investigate the V-Type ATPase complex within the context of pathways, the Oxidative Phosphorylation pathway (*bmy00190*) of *B. malayi* was visualised using the PathView program[258], which overlays the $\log_2$ fold-changes from EdgeR onto publicly available pathways, such as those hosted by the Kyoto Encyclopaedia for Genes and Genomes (KEGG[255]).

The V-Type ATPase complex as a whole was highly downregulated (**Figure 5-5**). Of the 15 components that make up the complex, 10 showed statistically significant downregulation (FDR < 0.05) by EdgeR analysis (**Figure 5-5**, genes with red outline), with an additional 3 genes approaching statistical significance (FDR < 0.1, **Figure 5-5**, genes with dark orange outline). This observation corroborates results observed via GO term enrichment analysis. Interestingly, the remainder of *B. malayi*'s oxidative phosphorylation pathway showed limited to no differential expression. Specifically, bacterial complexes 1 through to 4, responsible for generating a proton gradient, and Complex 5, or the ATP synthase complex, showed limited to no differential expression.

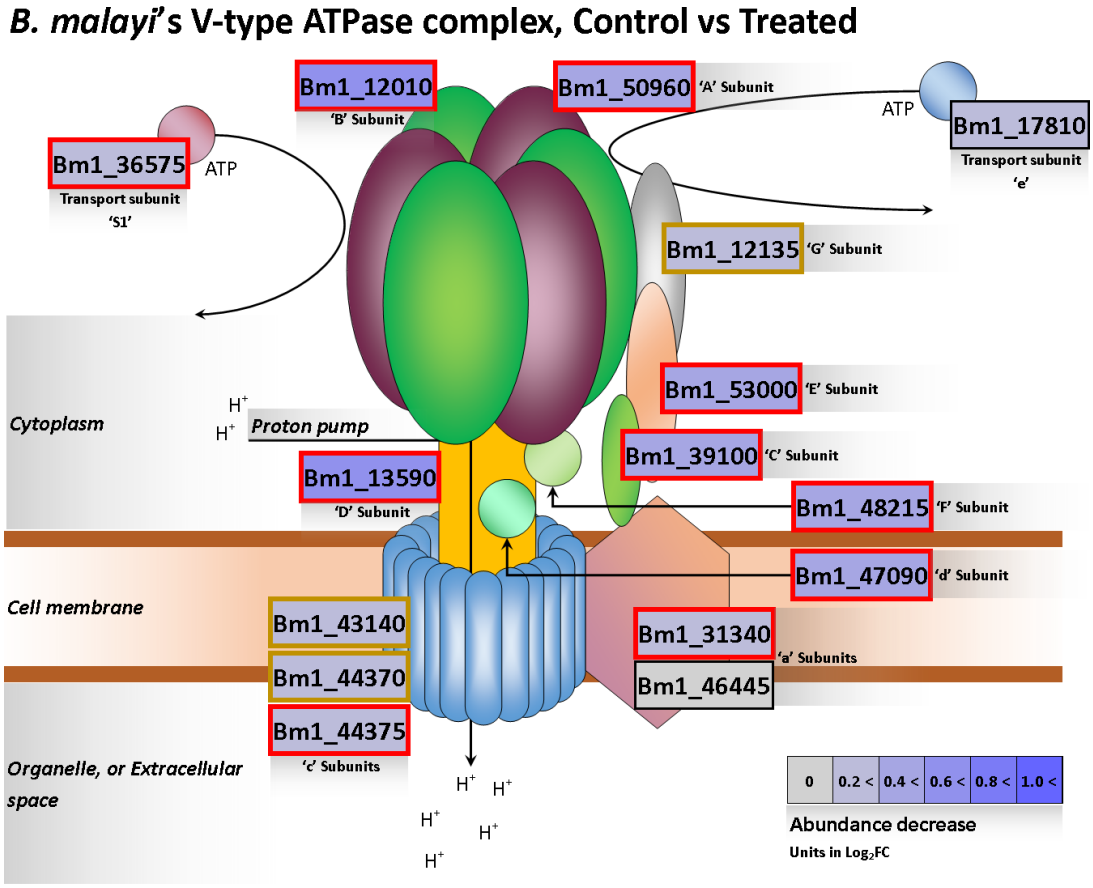


***Figure 5-5*: Comparison of transcriptional profile of B. malayi's V-Type ATPase complex after treatment and wBm removal.** *Individual gene IDs are shown in boxes, of which those with a red or dark orange outline show statistical significance (FDR < 0.05), or approaching statistical significance (FDR < 0.1) respectively. All genes are coloured in according to their expression fold-change. Note how 14 of the components of this pathway show downregulation within the treated group as compared to the untreated group, and that 10 of these components show statistical significance (FDR < 0.05), with an additional 3 approaching statistical significance (FDR < 0.1).*

5.2.9 Reverse-Transcription qPCR confirmation of observations

To confirm the validity of differential expression results, Reverse-Transcription quantification PCR (RT-qPCR) was performed on a total of 6 genes, utilising the same rRNA-depleted samples sent for sequencing to confirm accuracy of results. A 'test' set of 4 genes were identified that had a combination of: a minimum absolute $\log_2$ fold-change of 1, minimum raw read counts in both treated- and untreated samples of 1,000, at least one gene that showed up- or down-regulation, and an FDR < 0.05. These 4 genes encoded 2 kinases (*Bm1_52030*, *Bm1_55750*), 1 PAN domain containing protein (*Bm1_20410*), and 1 organic anion transporter (*Bm1_30620*). These were accompanied by an additional 2 'housekeeping' genes to act as references, chosen based on their protein product (actin and DNA polymerase, *Bm1_17980* and *Bm1_49670* respectively), as well as an EdgeR-calculated fold-change value at or close to 0. For the analysis of the chosen 6 targets, we utilised the Pfaffl method[307] to normalise the variance observed in qPCR amplification efficiencies. This is an

important metric for qPCR experiments, which measures the ability of amplicons to double with every thermocycle, and can vary significantly depending on the primers or reaction mixtures used. For comparative analysis of multiple targets, reaction efficiencies should be as close to each other as possible, or otherwise normalised mathematically[307].

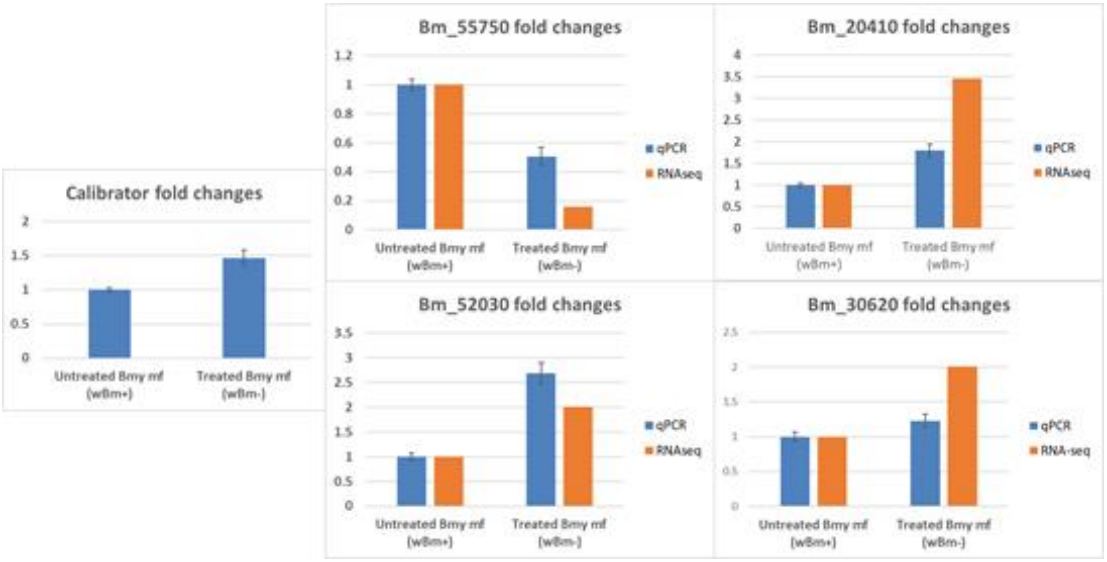


***Figure 5-6**: **Fold-change results from qPCR analysis of the 2 calibrator and 4 test genes, the latter compared with results from EdgeR differential expression analysis.** Note how qPCR results for all 4 genes show a generally similar trend to EdgeR results.*

All 4 test genes showed good agreement in directionality between treated- and untreated samples when compared to the results from EdgeR (**Figure 5-6**), strengthening the validity of differential expression results. As RT-qPCR and RNA-sequencing rely on different techniques and reactions, identical results in regards to fold-change between the two experiments was not believed to be likely and indeed not observed. The 2 calibrator samples experienced an approximate 1.5 fold-change increase within the treated samples based on qPCR. This is possibly due to differences in starting material concentration for the reverse-transcription reaction, as treated samples yielded slightly higher concentrations of RNA after Terminator Exonuclease treatment as compared to untreated samples (**Table 5-3**).

5.3 Discussion

Here we confirm that the filarial nematode *B. malayi* requires the presence of its *Wolbachia* endosymbiont for successful midgut penetration of the insect vector, mirroring past results from work done by Cook *et al.* (unpublished work). While apparently essential for midgut penetration, as shown by comparison of blood feeds of treated- or untreated mf to mosquitoes, *w*Bm's presence or absence did not show a statistically significant effect on *B. malayi*'s ability to develop from mf to infective L3 stages. However, it was observed that

mosquitoes injected with untreated mf had a reduced L3 recovery rate when compared to mosquitoes fed with untreated mf. As such, the lack of statistical significance in the injection experiments of *B. malayi* directly into the mosquito may be influenced by the overall reduced rate of development that is observed within the injected group.

A potential explanation for this reduced development rate, is that the act of gut translocation by fed mf directly affects the success of the exsheathment process, which may be a prerequisite for successful development in the mosquito. As injection directly bypasses the need to traverse the mosquito gut wall, injected mf may not be able to exsheath successfully, thus affecting their ability to develop within the mosquito. Further experiments may need to be conducted, such as evaluating whether the mode of mf invasion of the mosquito influences the success of subsequent development, or whether repeating this experiment with a larger sample size, to account for reduced L3 recovery rate in the injected group, may show statistical significance.

With regards to transcriptomic differences between the tetracycline- treated and untreated mf groups, notable are the 4 out of 5 genes that encode a protein product with chitinase activity that show statistically significant downregulation within the *Wolbachia*-depleted group. Chitin is the primary carbohydrate component of many insects' connective tissues, functioning as a scaffold material to support the cuticle of multiple insect tissues, such as the epidermis, trachea, as well as the peritrophic matrix found in blood-feeding insects, such as the mosquito vector of *B. malayi*[329,333,343]. Stage-specific expression of chitinase in mf may have a role in allowing penetration of the gut lining of the mosquito vector, and subsequent migration to mosquito flight muscles to continue their development. Chitinase transcripts are significantly downregulated in *Wolbachia*-depleted nematodes and may thus prevent the nematode from successfully penetrating the midgut of the vector.

Chitin is also known to be the primary support molecule within many mf sheaths[270,341]. In the case of *B. malayi*, mf are released from adult nematodes encased within this chitinous membrane, and they do not shed this sheath until during the process of penetrating a susceptible mosquito vector's gut wall[11,12,326]. Wu *et al.* has previously shown that *B. malayi* chitinase is located in a mf-specific structure called the Inner Body[341], and is excreted from this structure to other tissues of the nematode, and into the extracellular space between the nematode and the sheath upon mf entry into the midgut of the mosquito[341]. They have also shown that the activity of this chitinase can be inhibited in the presence of Allosamidin, and that the rate of mf exsheathment has a negative correlation with the concentration of

Allosamidin[341]. Prior to this, microscopy studies had reported the presence of physiological structures on the heads of mf, and had postulated that these may play a role in aiding invasion of the insect vector[12]. It may be possible that exsheathment via use of chitinase is a prerequisite for successful midgut penetration, either via these physical structures or the release of chitinase to break down the mosquito midguts.

Separately, the presence of statistically significant terms related to the V-type ATPase, and the downregulation of the complex as a whole, is of interest. The V-type ATPase is a proton pump located within intracellular membranes, and is believed to be widespread in many eukaryotic cells[337,344]. It is perhaps most well-known for its role in the acidification of intracellular compartments, or vacuoles. This acidification serves a multitude of roles within organisms, as reviewed by Beyenbach *et al.*[337]. Examples of the roles of acidification are in endocytosis and processing of raw materials for biosynthetic pathways in lysosomes, or pro-peptide processing for release into the extracellular space[337]. In more complex cases, the proton-pump action of V-type ATPases also plays a role in the control of receptor trafficking or synaptic release in nervous systems via use of Na^+, Ca^{2+}, or Cd^{2+} / H^+ antiporters[337,345]. Acidification of the *B. malayi* inner body, or some other synaptic signalling mechanism, may be required for the release of matured chitinase enzymes to allow nematode exsheathment and/or vector gut wall penetration, with the downregulation of transcripts for the V-type ATPase thus causing an inability for this release to take place.

As to why this effect is observed after *Wolbachia* depletion, *Wolbachia* endosymbionts of nematodes are known to be localised within vacuoles derived from the host's lipid membrane[29,58], and maintain an array of surface proteins that may play a role in interactions with their nematode hosts[111,271]. For example, the *Wolbachia* surface protein has been shown to associate itself with several components of the nematode glycolytic machinery[152,276], and there is evidence that *Wolbachia* is capable of co-opting host cytoskeleton microtubules to aid in its migration around host tissues[152,271]. Indeed, within the list of enriched GO terms within the downregulated gene set, we see several terms related to actin filaments, of which *Wolbachia* are known to use to aid its migration between host cells. It may be possible that the depletion of the *Wolbachia* endosymbiont in these mf has resulted in a number of signalling changes, resulting in the downregulation of transcripts for synthesising and maintaining cytoskeleton tubules, as well as proton/vacuolar pumps.

The presence of *w*Bm may also play a role in a basal level of signalling that helps to stimulate nematode larval development, with several GO terms related to nematode larval

development showing downregulation in the treated group of nematodes. While such results are not convincingly reflected via the direct injection of mf into the *A. aegypti* vector, possibly due to too small a sample size, it does mirror what is seen within *O. volvulus* nematodes[86], and their ability to develop to the infective L3 stages[86].

In conclusion, this transcriptomics study has shown that *B. malayi* mf undergo targeted transcriptional changes after depletion of its *Wolbachia* endosymbiont. These changes centre on its ability to synthesise a stage-specific chitinase, and possibly process it for secretion to the extracellular environment. This will likely have profound effects on the nematode with regards to its ability to successfully invade and develop within the intermediate vector host. Such observations begin to challenge the previously held paradigm that *Wolbachia*'s presence is only necessary from L3-stage nematodes and onwards, and demonstrates that *Wolbachia* plays a role in other key life-cycle stages of the nematode host.

5.4 Methods

5.4.1 Animal model and parasites used

The life cycle of *Brugia malayi* was maintained at LSTM for experimental use. In summary, infected jirds were first anaesthetised with isofluorine and subjected to peritoneal washes with RPMI 1640 media (ThermoFisher Scientific). These washes were then extracted via peritoneal catheterisation to harvest mf. These were then purified via use of PD10 column size exclusion chromatography (Amersham), counted via microscopy and then mixed with human blood to a final concentration of 15-20,000 mf/ml. They were then fed to female *Aedes aegypti* mosquitoes via an artificial membrane feeder (Hemotek). Blood-fed mosquitoes were then reared for 14 days with daily sugar-water feeds to allow the *B. malayi* larvae to develop to infective L3-stage nematodes. At day 14, L3-stage nematodes were collected via crushing of the infected mosquitoes and subsequent concentration via use of a Baermann's apparatus and RPMI media. The L3 nematodes can then be used to continue the life cycle via injection into additional jirds.

For this study, 400 infective-stage L3s were injected into the peritoneal cavity of 9 jirds, aged minimum 4 months. Animals were then left for a minimum of 16 weeks to allow *B. malayi* nematodes to develop into reproductive adults, and begin production of mf. Subsequent to this 16 week incubation phase, 6 jirds were given Tetracycline in their drinking water, at a 0.5% concentration daily, with an additional 3 jirds being negative controls. This dosing regimen continued for a total of 6 weeks.

Upon completion of the dose regimen, all jirds were culled via asphyxiation and dissected to reveal the peritoneal cavity. Several washes of all jirds' peritoneal cavities were then conducted using RPMI media to retrieve any mf present, which were then purified via PD10 column size exclusion chromatography and counted. Mf were then separated into aliquots of 200,000 nematodes (treated group) and 500,000 nematodes (untreated group) and stored in RNA-Later (Invitrogen). Samples were kept as separate biological replicates where possible.

5.4.2 Simultaneous RNA and DNA extraction

Mf in RNA-Later were first diluted via adding an equal volume of phosphate-buffered saline to the sample, then centriuged at 2,000x g at 4°C for 10 minutes to pellet mf. The supernatant was removed, before 1ml of TriZol Plus reagent (Invitrogen) was added. Samples were then transferred to separate sterile 2ml screw cap tubes containing ceramic beads of different sizes (1.4 and 2.8mm) (CKMix, Bertin). Batches were homogenized at 6,000rpm (Minilys, Bertin Instruments) for 4 x 30 seconds, cooling on wet ice for 30 seconds in between.

200µl molecular grade chloroform (Invitrogen) was then added for every 1ml of TRIzol reagent, and mixed by vigorously shaking the tube for 15 seconds, followed by incubation at room temperature for 5 minutes. The sample was then centrifuged at 12,000x g for 15 minutes at 4°C. The upper aqueous phase was removed and transferred to a nuclease-free 1.5ml tube. An equal volume of ice cold 70% ethanol was added to each sample and vortexed for 2 - 3 seconds. The tube was inverted several times to disperse any precipitate.

Subsequent binding, washing and elution of RNA from the sample was then conducted using Exiqon's miRCURY™ RNA Isolation Kit – Cell & Plant, in accordance to manufacturer's instructions. gDNA was also removed via an on-column treatment using Baseline-ZERO™ DNase (Epicentre). Final RNA was eluted in 2 x 50µl molecular grade water (Invitrogen).

The remaining homogenate was then used for DNA extraction, following TriZol manufacturer's protocol via ethanol precipitation.

5.4.3 qPCR analysis of *Wolbachia* populations

Extracted gDNA from all samples was amplified with primers that target the *Wolbachia wsp* and *B. malayi gst* genes. Amplification of the *wsp* gene utilised a final concentration of 0.3µM for each primer, and 3.0µM of $MgCl_2$, with annealing conducted at 60°C. Amplification of the *gst* gene utilised a final concentration of 0.35 µM for each primer, and 1.5µM of $MgCl_2$, with annealing conducted at 55°C. All tests were carried out in triplicate, within a total volume of 20µl and included 2µl of DNA from each sample, and 10µl of QuantiTect SYBR Green PCR kit

master mix (Qiagen). For both *wsp* and *gst*, a standard curve was also generated in triplicate. All qPCR reactions were conducted using a CFX 384 Real Time PCR detection system (BioRad), with all reactions heated for an initial 15 minutes at 95°C, before being subjected to 40 cycles of 15 seconds at 94°C, 30 seconds at the specified annealing temperature, and 30 seconds at 72°C.

5.4.4 Mosquito maintenance, feeds and injection of microfilariae

Aedes aegypti female mosquitoes were reared from eggs obtained from the National Institute of Health, Filariasis Research Reagent Resource Center (FR3). Adult mosquitoes between 3 to 5 days old were sucrose-starved (between 18 to 20 hours), before being fed on human blood inoculated with mf (200 mf per 20µl of blood), collected from gerbils as described previously. Feeding was performed using the Hemotek feeding apparatus (Discovery Workshops, UK) covered with parafilm.

Injection of mosquito thoraxes with *B. malayi* mf followed a protocol as described by Erickson *et al.*[346]. Briefly, mf were isolated from blood samples via syringe tip filtration devices fitted with 5µm membranes (Milipore Isopore TMTP). Mf were washed in up to 2ml of *Aedes* saline solution in conical vials, before being spun at 1,000 RPM for 10 minutes at 4°C and removal of supernatant to concentrate parasites. This concentrated parasite saline solution was transferred to a microscope slide, and mf were loaded into finely pulled glass capillary needles for injection into mosquitoes. A dissecting microscope was used to monitor this loading process, and ensure between 10-20 mf were loaded per needle. Mosquitoes were cold anesthetised for 3 minutes at -20°C immediately prior to injection, and were injected with between 0.5 to 1µl of *Aedes* saline containing the loaded mf, into a membranous cuticle area on the side of the mesothorax.

Both blood-fed and injected mosquitoes were then transferred to a fresh cage and kept for 14 days post-feeding, before being dissected in phosphate-buffered saline to count the number of L3 stage parasites.

Subsequent to dissection and L3 recovery from mosquitoes, a type-I two-way ANOVA test was performed in the programming language R to account for the difference in sample sizes between the bloodfed and injected conditions. This test was then followed by Tukey-Kramer multiple pairwise comparisons tests to identify which conditions showed statistical significance. The results of this were then plotted in R using the ggplot2 package. Code for significance tests is as follows:

```
mf.test.anova <- aov(Counts ~ Injected * Treated, data = mf.experiment)
```

TukeyHSD(mf.test.anova)

5.4.5 Illumina RNA Library Preparation and Sequencing

RNA samples were assayed for total RNA concentration and purity via the use of a Nanodrop spectrophotometer (ThermoFisher Scientific), and subsequent quality analysis using a 2100 Agilent Bioanalyzer machine system (Agilent, Total RNA Nano chip), with all samples having a minimum RNA-Integrity score of 8, or having clearly defined peaks for both 18S and 28S rRNA. In order to remove excess ribosomal RNA species from both *Wolbachia* and *B. malayi*, total RNA was treated with Terminator 5'Phosphate-Dependent Exonuclease (Epicentre), in accordance to the manufacturer's instructions. The samples were then cleaned for chemical and protein impurities via use of a commercial purification kit (Zymo, RNA Clean & Concentrator-5), following manufacturer's instructions.

rRNA-depleted samples were then quantified via use of a Qbit 3.0 machine (Invitrogen), before being sent to the University of Liverpool's Centre for Genomic Resources for RNA sequencing via use of an Illumina HiSeq 4000, and the NEBNext Ultra Directional RNA library preparation kit (New England Biolabs). Two separate paired-end sequencing runs (2x 150 bp) were performed for each sample. Files for all data were then converted from raw basecall to fastq using CASAVA 1.8 (Illumina). The raw Fastq data files were trimmed for the presence of Illumina-specific adapter sequences using Cutadapt (v1.2.1)[243] with the option *–O 3* and further trimmed for quality using Sickle (ver.1.200)[244], with a minimum quality window score of 20, and reads less than 20 bases were discarded. Sequencing quality scores were assessed using FASTQC v0.9.2.

5.4.6 Mapping and Quantifying RNA-sequence Data

RNA-sequence reads from processed fastq files were aligned to the genome of *B. malayi* (GenBank ID GCF_000002995.3, ASM299v2) using Subread-Aligner's Subjunc program (ver. 1.5.0)[188] and the options *--multiMapping*, *-B 2*, giving a Binary Alignment/Map (BAM) output file. The BAM files were then separated into aligned and unaligned reads using Samtools[221] (ver. 0.1.19), with one BAM file containing aligned RNA-sequence reads mapped to *B. malayi*.

The aligned RNA-sequence data from *B. malayi* were quantified against its genome annotation held on the NCBI database. This quantification used the program FeatureCounts[252] (version 1.5.0-p3) to obtain read counts that were then subsequently used in differential expression analysis.

5.4.7 Differential Expression Analysis and Pathway mapping

Differential expression was conducted on the *B. malayi* read counts using the program EdgeR (version 3.10.5) and an online interactive wrapper, Degust (version 3.2.0)[127,259], using as input the read counts obtained from FeatureCounts (version 1.6.2). Differential expression utilised a pairwise comparison, with untreated nematodes acting as the reference point against nematodes treated with tetracycline for 6 weeks. Genes were deemed as statistically significantly differentially expressed using a threshold of FDR < 0.05 – a standard setting in EdgeR. Differential expression results were visualised in a volcano plot using ggplot2[347].

Gene Ontology analysis was conducted via separating the list of significantly differentially expressed genes into up- or downregulated genes. These two sets were then used as a 'test' set, and were analysed against the entirety of *B. malayi*'s genome as a 'background' set using Blast2GO's Fisher's Exact Test functionality[182], using an FDR cutoff of < 0.05 for significance.

For pathway analysis of the *B. malayi* transcriptome, gene IDs were converted using BioDBnet's db2db online resource[285], from GenBank nucleotide accession to KEGG gene ID. These converted gene IDs, and their associated fold-change results from EdgeR analysis were used as input into the PathView program[258] (ver. 1.22.0), using the options *kegg.native = T, node.sum = mean*. In addition, the fold-change results from EdgeR was also used as input for the Generally Applicable Gene set Enrichment program[135] (ver. 2.32.0), using the options *gsets = kegg.gs, same.dir = TRUE, saaTest = gs.KSTest* for analysis of up- or down-regulated pathways.

5.4.8 Reverse-Transcription qPCR validation analysis

To confirm results of differential expression analysis, the DNAse-treated, rRNA-depleted samples that were prepared for sequencing were used as template for cDNA synthesis, using the SuperScript IV First Strand Synthesis System (Invitrogen) following manufacturer's instructions.

Four genes were selected for analysis were based on a minimum raw read count > 1,000, a minimum absolute $\log_2$ FC > 1, and an FDR < 0.05. The 2 genes selected to act as calibrator genes were selected based on their predicted housekeeping functions as actin and RNA polymerase genes, as well as constitutive expression across the two arms of the experiment based on RNA-seq data. Up to 3 primer pairs were designed for each gene, with primer efficiencies validated using cDNA of *B. malayi* mf generated separately, tested in triplicate. Primers with the highest efficiencies were selected for confirmation of differential expression results using test cDNA.

Amplification of all 6 genes utilised a final concentration of 0.3µM for each primer. All tests were carried out in triplicate, within a total volume of 20µl and included 2µl of DNA from each sample, and 10µl of QuantiTect SYBR Green PCR kit master mix (Qiagen). A standard curve was also generated for all 6 genes in triplicate. All qPCR reactions were conducted using a CFX 384 Real Time PCR detection system (BioRad), with all reactions heated for an initial 15 minutes at 95°C, before subjected to 40 cycles of 30 seconds at 94°C, 30 seconds at 60°C, and 30 seconds at 72°C.

Chapter 6 General discussion and future perspectives

Advances in sequencing and computing technologies have allowed for a better understanding of the biological basis of *Wolbachia*-filarial nematode relationships, particularly during adult stages where *Wolbachia* depletion results in a halt in embryogenesis and macrofilaricidal effects[169,288,348,349]. Despite this, comparatively little is known about the basic biology and symbiotic relationship between these two organisms during early life cycle developmental phases of the nematode host, such as the L3-L4 developmental stages, and microfilariae-vector transmission stages. Additionally, the genome annotations for the main experimental organisms for study, *Brugia malayi*, and its *Wolbachia* endosymbiont (*w*Bm), are over a decade old as of 2019[111,112], and in need of updates with new knowledge and tools. The work described in this thesis has attempted to address these gaps in knowledge via utilisation of RNA-sequencing technologies to investigate the transcriptome of *B. malayi* and *w*Bm across different developmental life cycles. This has resulted in the reannotation and partial characterisation of new genes in *w*Bm's genome (**Chapter 2 and Chapter 4**), as well as an investigation of the mutualistic relationship the two organisms share during key life cycle stages which have received comparatively little attention in existing literature (**Chapter 3 and Chapter 5**).

6.1 New genes of *w*Bm linked to interactions with *B. malayi* host

Utilising RNA-sequencing data and alignments to the *w*Bm genome, this study successfully identified a total of 21 protein-coding genes and 3 functional RNA genes in *w*Bm (**Chapter 2**). Within these 21 identified protein-coding genes were components of the Type IV secretion system (*VirB2, VirB7*), responsible for the transport of DNA and/or proteins from bacteria to a target[194,225]. Additionally, one component of the Sec-secretion system was also identified (*SecE*). This secretion system is responsible for insertion of proteins from the bacterial cytosol into the membrane, or exported extracellularly[196], where they could be used as a signalling mechanism to interact with the nematode host. Whilst the majority of components for both systems were previously identified by Foster *et al.*[111], these newly identified components were noticeably absent, yet perform key functions in their respective transport systems. Their annotation and evidence of expression by RNA-sequencing, provides clear confirmation that both systems are active and functional and helps to improve our understanding *Wolbachia* interactions with the nematode host.

Following on from this, 2 genes that had no identifiable function were also identified as of potential significant interest in the interactions between *w*Bm and the nematode host. One of these genes (*wBmNew0001*) appeared to be unique to the *Wolbachia* genus, with all

identifiable homologues noted to maintain well-conserved transmembrane domains. This contrasted with a predicted intracellular C-terminus tail that showed noticeable sequence variation amongst identified homologues, with this variation demarcated by supergroup origin. As *Wolbachia* in general are known to be localised within host-derived vacuoles[29,30], it is possible that the C-terminus of this gene's protein product is localised within the host cytoplasm. These two factors, supergroup-specific variability and localisation, implies that the protein product of this gene may have host-specific interactions, possibly for transport or signalling. Whilst previous studies have focused on what genes are, or are not, encoded between the genomes of sequenced *Wolbachia*[111,118,350,351], little has been done to investigate the variation of gene homologues between different supergroups of *Wolbachia*. Expanding from this, it is possible that the *Wolbachia* genus maintains several genes similar to *wBmNew0001* that have protein domains with sequence variation that can be demarcated by supergroup. Such genes may play conserved, supergroup-specific roles in the relationship that individual *Wolbachia* have with their hosts, whether it be arthropod or nematode.

The second gene of interest was identified by both transcriptomic and proteomic analysis (**Chapter 4**). This gene had conserved residues that mapped partially to well-conserved microtubule-binding proteins of the *Eg-5* spindle pole kinesin motor proteins, responsible for mitotic spindle formation and chromosome segregation[301,302]. Members of the *Wolbachia* genus are known to lack genes for canonical bacterial motility, such as pili or fimbriae, and had previously been believed to subvert host actin filaments to achieve intracellular mobility[111]. As such, the presence of genes that could have microtubule binding function is significant in our understanding of *Wolbachia* migration and localisation within host cells. Yet attempts to identify any full-length homologues in other *Wolbachia* genomes was unsuccessful, with the closely-related *Wolbachia* of *Wuchereria bancrofti* maintaining a homologue that is interrupted by both stop-codons and frame-shifts. This points to the gene being unique to *w*Bm, and while there is a possibility of sequencing errors in the *Wolbachia* of *Wuchereria bancrofti* genome, this gene may have a unique function in the *w*Bm-nematode symbiosis.

6.2 Regulation of gene expression by *w*Bm is complex and supports rapid bacterial proliferation

Following on from the reannotation of *w*Bm's genome via RNA-sequencing, 5 instances of non-model translational events, including Stop-Codon Read-through (SCR) or Programmed Ribosomal Frame-shifting (PRF) were also identified in *w*Bm (**Chapter 2**). These events are characterised by the ribosome ignoring stop codons or changing reading frames during

translation of an mRNA transcript, respectively. These processes appear to be widespread across all taxa, being well-known in viruses[205,234], known to occur within prokaryotes[235,236,352], and more recently have been observed in multicellular eukaryotes[208,209]. Such events are believed to play roles in gene regulatory mechanisms in response to external stressors such as environmental changes[235,236]. The observations of highly expressed genes in *w*Bm that appear to undergo these processes is the first such indication that these events may occur within members of the *Wolbachia* genus and may be wide-spread amongst *Wolbachia* in general. Considering the wide host range that *Wolbachia* is known to infect, as well as the various different life cycle stages these hosts proceed through, appropriate identification of such genes will likely be important for both new and existing *Wolbachia* genomes. Such genes that undergo non-model translational events may offer glimpses of unique biology that could play roles in the *Wolbachia*-host relationship. In *w*Bm for instance, the 5 genes that undergo such non-model translational processes encode for transmembrane transport genes, as well as protein hydrolase genes. These genes may be translated only at key life cycle stages of the host, either to allow for supplementation of metabolites to the host, or possibly utilised by *w*Bm itself to co-opt host amino acids and supplement its own needs, as it is known to lack *de-novo* amino acid biosynthesis pathways[111].

Following on from this translational control, previous literature has noted a lack of transcriptional regulators within the genomes of *Wolbachia* in general, and had postulated that most of the *Wolbachia* genome would be expressed constitutively[111]. This has been observed in this study, as only 8 genes could be identified as being uniquely expressed in the transcriptome of *w*Bm in L3-L4 developing nematodes, when compared to the transcriptome of *w*Bm in microfilariae (**Chapter 3 and Chapter 5**). Yet despite this constitutive expression, this study was able to observe coordinated, yet rapid, transcriptional control in *w*Bm in relation to its environment. The best example of this is the observed transcriptional control over carbon metabolism during early stages of the nematode host's L3-L4 development, with an observed preference for the glycolysis pathway over the tricarboxylic acid (TCA) cycle in *w*Bm (**Chapter 3**). This preference for glycolysis has been observed to occur in cancer cells even within oxygen-rich environments, and is known as the Warburg effect[272], or more generally as aerobic glycolysis in other cell types. As glycolysis has a high rate of flux and many of the pathway's intermediates have use in biomass-generating pathways, this preference has been hypothesised to fuel rapid cellular proliferation[272,274]. This may be the basis for the observed rapid growth of *w*Bm in the immediate period following infection of the mammalian host. After the first week a reduction in *w*Bm cellular replication leads to a

switch in transcriptional control with *w*Bm showing a preference for the TCA cycle over the glycolytic pathway. This transcriptional control of carbon metabolism indicates that *w*Bm is capable of sensing both local conditions in its host's environment and can respond accordingly.

It should be noted that this preference for glycolysis even in oxygen-rich environments and where glucose is plentiful, produces large amounts of waste metabolites via fermentation. For instance, yeasts and bacteria are also known to preferentially utilise glycolysis in plentiful, oxygen-rich environments to fuel cellular proliferation leading to the production of ethanol. In this system, the process is referred to as the Crabtree effect[273]. Aerobic glycolysis has also featured during the erythrocytic cycle of *Plasmodium* parasites as well[274,353]. Furthermore, healthy human immune cells, such as T-cells and macrophages, have been observed to utilise aerobic glycolysis in order to proliferate and produce host defence factors in response to an infection[353–355], whilst producing large amounts of lactate as a waste product[353,355]. This lactate is the same waste product produced by cancer cells via the Warburg effect, and is often reused as input for the TCA cycle for energy production[272,355]. As such, aerobic glycolysis is a conserved mechanism across many different taxa, although this has yet to be reported within *Wolbachia*. Such observations may be exploitable in encouraging *Wolbachia* growth or infection, specifically for the artificial infection of arthropods for pathogen blocking strategies[48–50,356], although this utilisation of glycolysis for cellular proliferation needs to be confirmed in *Wolbachia* of arthropods.

6.3 The variable relationship of *w*Bm and *Brugia malayi* across developmental stages

Whilst the transcriptomic data indicates a preference for aerobic glycolysis in *w*Bm to fuel cellular proliferation, uniquely in *w*Bm is an inability to utilise glucose as part of the glycolysis pathway, and thus produce waste products via fermentation. This has been predicted due to a lack of genes that allow for the uptake and phosphorylation of extracellular glucose[23,111] (such as glucokinase or glucose-1-phosphatases). Instead *w*Bm has been predicted to utilise gluconeogenesis[23] starting from pyruvate, which may be provisioned by the nematode. Supporting evidence for this includes the presence of several genes in *w*Bm that act solely in the gluconeogenic direction, or are unusual in having both gluconeogenic and glycolytic functionality, such as fructose 1,6 bisphosphatase, or pyruvate phosphate dikinase respectively. Additional evidence also includes the co-localisation of host-derived glycolysis enzymes to the *w*Bm-containing vacuole[152], observed previously in adult nematodes. This association appears critical for the fitness of the *w*Bm endosymbiont, as RNA-interference or

drug-based treatment to inhibit nematode glycolysis pathway genes results in a subsequent reduction in *w*Bm population[276,277], and supplementation of nematode maintenance media with pyruvate showed either an increase in *w*Bm population in untreated nematodes, or at least a partial reversal in treated nematodes[277]. Whilst these observations have been made in adult nematodes, such observations have yet to be made in L3-L4 nematodes before this study. These observations indicate that the nematode may actively encourage *w*Bm population growth during multiple developmental life cycle stages, including immediately after nematode infection of the mammalian host. During later stages however, the nematode begins a careful balancing act of maintaining a sufficient *w*Bm population to fuel its own growth, without causing a deleterious load of *w*Bm via the use of the autophagy pathway. *B. malayi* utilisation of the autophagy pathway is in line with previous observations, which show an abundance of the protein ATG8a, a biomarker for autophagy activation, in L4-stage nematodes as well as in adults, but not in L3-stage nematodes[266].

Outside of this control over carbon metabolism, upregulation of several other pathways was observed in *w*Bm from Day 7 to Day 14, with prominent examples being *de-novo* haem and nucleotide biosynthetic pathways (**Chapter 3**). These have long been suspected in playing a role in the symbiotic relationship between *w*Bm and *B. malayi* due to comparative genomics studies identifying incomplete, or absent, pathways in the nematode, yet intact versions within the endosymbiont[23,111,112,247]. This constitutive upregulation complements the transcriptome of *B. malayi*, which appears principally focused on its own growth during the 2 weeks of study. Following on from this, almost all components of *w*Bm's Type IV secretion system, including those newly identified as part of reannotation work (**Chapter 2**), were also constitutively upregulated across all 4 studied time-points.

Expanding observations to other life cycle stages, evidence for a potential role of *w*Bm in permitting the nematode host's successful infection of the vector host has been growing[85,86], yet the actual mechanism behind this process is not understood. The transcriptomics experiments performed as part of this thesis show that the *B. malayi* host undergoes targeted transcriptional changes in response to depletion of its *w*Bm endosymbiont (**Chapter 5**). Notable amongst these changes are fold-change reductions in transcripts encoding for multiple chitinase enzymes in *B. malayi*, which have been previously implicated in mf exsheathment and vector host midgut penetration[341], a phenotype observed in the blocking of *Wolbachia* depleted-mf transmission in the mosquito vector. In addition to these transcripts, this study also identified downregulation of the *B. malayi* V-type ATPase complex as a whole, which plays various roles in vacuole acidification and enabling the maturation of

enzymes, as well as autophagosome maturation and synaptic signalling[337,345]. When *w*Bm is present, upregulation of V-type ATPase components may be required to maintain *w*Bm's static populations within microfilariae. Additionally, these complexes may also have roles in processing and activating chitinase enzymes within the nematode inner body via pro-peptide cleavage. Thus, depletion of *w*Bm would result in reduced requirements of the V-type ATPase for *w*Bm population control, but also have a secondary effect of reducing the ability to process chitinase enzymes, and thus affecting the nematode's ability to exsheath and traverse the vector midgut. Whilst it is still unclear as to whether *w*Bm is, or is not, provisioning metabolites to microfilariae on account of their continued motility and viability following *Wolbachia* depletion, at the very minimum its presence appears to be required for successful transmission. This, and previous observations, highlight the diverse role of the *w*Bm endosymbiont across multiple developmental life cycle stages of the nematode, and opens new avenues for control of disease transmission in addition to therapeutic intervention for permanent sterilisation and macrofilaricidal cure.

It is interesting to note that transcripts for these chitinase enzymes were identified only during the microfilariae stages of *B. malayi* (**Chapter 5**) and could not be identified during the L3-L4 developmental transcriptome (**Chapter 3**). Such observations are similar to previous studies which identified chitinase as an mf stage-specific enzyme, and not identified in any other life cycle stage of *B. malayi*[341]. Other filarial nematode species that have a *Wolbachia* endosymbiont, such as *Onchocercha volvulus*, are also known to express stage-specific chitinases, although in this instance they are limited to the L3 developmental stage, just before infection of the final host[338,339]. The reasons for this stage-specific expression in *O. volvulus* nematodes, or if their expression has any link to the *Wolbachia* endosymbiont, has yet to be elucidated.

6.4 Future work

Transcriptome analysis of both *w*Bm and *B. malayi* has revealed further insights into the nature of their mutualistic association, yet simultaneously raised further questions for future research. For instance, whilst new genes have been discovered in *w*Bm, some of which have been deemed of interest due to the presence of key domains that could play a role in nematode-*Wolbachia* interactions, several of these require further experimental validation to define their functions. For instance, elucidation of the function of *wBmNew0001* could provide more information on the complex relationship that *Wolbachia* endosymbionts in general have with their nematode or arthropod hosts. It would also be of interest to identify whether homologues of *wBmNew0004* exist in other *Wolbachia* genomes. If no such

homologues are identifiable, it would be interesting to test whether this protein is capable of binding to microtubules as predicted, and determine why this protein's function appears to be unique to *w*Bm.

Similarly, whilst the observation that the glycolysis/gluconeogenesis pathways may play a key role in *w*Bm cellular proliferation, further validation of these observations via metabolomics experiments is warranted. This could utilise isotopic labelling of carbon (such as glucose incorporated with carbon-13), and subsequent monitoring of the accumulation of intermediates from carbon metabolism within the developing nematode and the *w*Bm endosymbiont coupled with mass and NMR spectroscopy analysis of metabolomic profiles. Alternatively, it would be interesting to investigate carbon metabolism in *Wolbachia* endosymbionts of other nematodes or arthropod hosts.

Questions also remain as to the exact role that *w*Bm plays in enabling the nematode host to successfully infect susceptible mosquito vectors. Due to the role that chitinases appear to play in mf exsheathment and vector transmission, directly inhibiting the expression of these chitinases, via enzymatic inhibitors or genetic manipulation would be of significant interest as a confirmatory study. Additionally, further work investigating the effects of *Wolbachia* depletion on the transcriptomes of other filarial nematodes of medical importance would be of interest to determine any filarial nematode commonality to mutualistic mechanisms. As previously mentioned, other parasitic filarial nematodes such as *O. volvulus* express chitinases at different developmental stages after infection of the arthropod vector[338,339]. While *Wolbachia* depletion in *O. volvulus* does not hamper the nematode's ability to infect the vector, it does result in reduced nematode development[86], but again the exact mechanisms underpinning this have yet to be elucidated. It would be interesting to see how *Wolbachia* depletion affects the transcriptomes of these other filarial nematodes and whether dysregulation of chitinase transcripts and/or V-type ATPase complexes are a common theme of *Wolbachia* depletion.

Finally, one of the main caveats of the work conducted in this thesis, is how all RNA-sequencing datasets generated in this study were overwhelmingly dominated by *B. malayi* transcripts, and that endosymbiont transcripts thus suffered from a reduced read depth. This compromises our ability to rigourously define differentially expressed pathways or genes within the *w*Bm transcriptome. As such, further work performing repeat RNA-sequencing experiments with the same model, as well as newly-released technology and techniques[138,193], would likely help to increase read depth of endosymbiont transcripts, and

enable more confident detection of differentially expressed genes. Expanding these techniques to investigate other life cycle stages of interest would also be interesting to further understand *w*Bm's diverse contributions to the biology of filarial nematodes.

6.5 Concluding remarks

Much still remains to be discovered of the mutualistic relationship shared between *Wolbachia* and their nematode hosts, particularly during nematode life cycle stages outside of reproductively active adults. The utilisation of RNA-sequencing to investigate poorly studied life cycle stages of *B. malayi* has resulted in a better understanding of the basic biology that underpins this relationship, and begins to fill critical gaps in knowledge, and expands the variety of roles that *w*Bm plays in *B. malayi* across the nematode life cycle. This has identified *Wolbachia* as not only a target for therapeutic treatment for permanent sterility or macrofilaricidal activity, but also as a means to block disease transmission. These discoveries have the potential to influence future research into the interactions that *Wolbachia* have with their nematode hosts in general, as well as for drug administration programmes, in pursuit of the global elimination of these debilitating group of diseases.

Appendix 1 New annotation of *w*Bm

In the interests of space and paper, data for this appendix section has been uploaded to the online repository, https://figshare.com/#. Links for each item have been generated and provided for browsing.

General Feature Format file with coding genes of *w*Bm
https://figshare.com/s/95bb9b695bbc25971bd9

A 1-1*: General Feature Format file containing both newly identified and existing genes of* w*Bm from Foster et al. 2005*[111]*.*

General Feature Format file with pseudogenes of *w*Bm
https://figshare.com/s/dca7edb1bdf49656100e

A 1-2 *General Feature Format file containing both newly identified and existing pseudogenes of* w*Bm from Foster et al. 2005*[111]*.*

Functional annotation of coding genes of *w*Bm
https://figshare.com/s/fd49b4b6eff7ff7239f6

A 1-3: *Excel table containing all coding genes of* w*Bm, and their predicted function.*

Functional annotation of pseudogenes of *w*Bm, accession
https://figshare.com/s/0fb3b8dc25fd22e115c8

A 1-4: *Excel table containing all pseudogenes of* w*Bm, and their predicted function.*

Full alignment of gene *wBmNew0004*
https://figshare.com/s/63e1514fff667e448c11

A 1-5: *Figure representing the alignment of gene* wBmNew0004 *with kinesin 5 family genes from a variety of higher eukaryotes.*

Single nucleotide variant call-file of *w*Bm
https://figshare.com/s/6863d4789effe522321c

A 1-6: *Variant Call File (VCF) format containing details on single nucleotide variants within the* w*Bm genome.*

Appendix 2 Differential expression results of the L3-L4 developmental transcriptome

In the interests of space and paper, some data for this appendix section has been uploaded to the online repository, https://figshare.com/#. Links for each item have been generated and provided for browsing.

	Total reads (millions)	% total reads mapping to Bmy genome	*wBm* reads (thousands)	% remaining reads mapping to wBm genome
Day 3-r1	28.95	94.89%	64.23	4.35%
Day 3-r2	25.13	92.52%	54.76	2.91%
Day 3-r3	29.52	95.51%	65.93	4.97%
Day 7-r1	16.13	90.66%	190.05	12.61%
Day 7-r2	42.62	89.17%	516.94	11.20%
Day 7-r3	33.72	93.13%	436.58	18.85%
Day 11-r1	29.28	94.94%	912.71	61.65%
Day 11-r2	25.75	93.16%	786.97	44.70%
Day 11-r3	31.23	95.08%	1,005.47	65.40%
Day 14-r1	22.48	94.77%	808.21	68.76%
Day 14-r2	28.44	93.47%	1,026.72	55.29%
Day 14-r3	29.29	94.96%	1,077.83	73.05%

A 2-1*: Table to illustrate the differences in read mapping efficiency between* B. malayi *and the* w*Bm endosymbiont, as well as how read mapping efficiencies generally increased as time of infection progressed.*

Differential expression analysis of *B. malayi* L3-L4 development using *CuffDiff* https://figshare.com/s/94528f60859c780ff5ba

A 2-2: *Excel table containing 6 sheets outlining the raw results from CuffDiff, when applied to the L3-L4 developmental transcriptome of B. malayi*

Differential expression analysis of *B. malayi* L3-L4 development using *DESeq2* https://figshare.com/s/027947ae243141b68c0d

A 2-3 *: Excel table containing 6 sheets outlining the raw results from* DESeq2*, when applied to the L3-L4 developmental transcriptome of* B. malayi

Differential expression analysis of *B. malayi* L3-L4 development using *edgeR* https://figshare.com/s/52f74af94fdb5c128382

A 2-4: *Excel table containing 6 sheets outlining the raw results from edgeR, when applied to the L3-L4 developmental transcriptome of* B. malayi

Differential expression analysis of *Wolbachia* of *B. malayi* L3-L4 development using *CuffDiff*
https://figshare.com/s/76b2c905bf208d415501

A 2-5: *Excel table containing 6 sheets outlining the raw results from* CuffDiff*, when applied to the* Wolbachia *endosymbiont during* B. malayi *L3-L4 development*

Differential expression analysis of *Wolbachia* of *B. malayi* L3-L4 development using *DESeq2*
https://figshare.com/s/14f6516ce954cca88961

A 2-6: *Excel table containing 6 sheets outlining the raw results from* DESeq2*, when applied to the* Wolbachia *endosymbiont during* B. malayi *L3-L4 development*

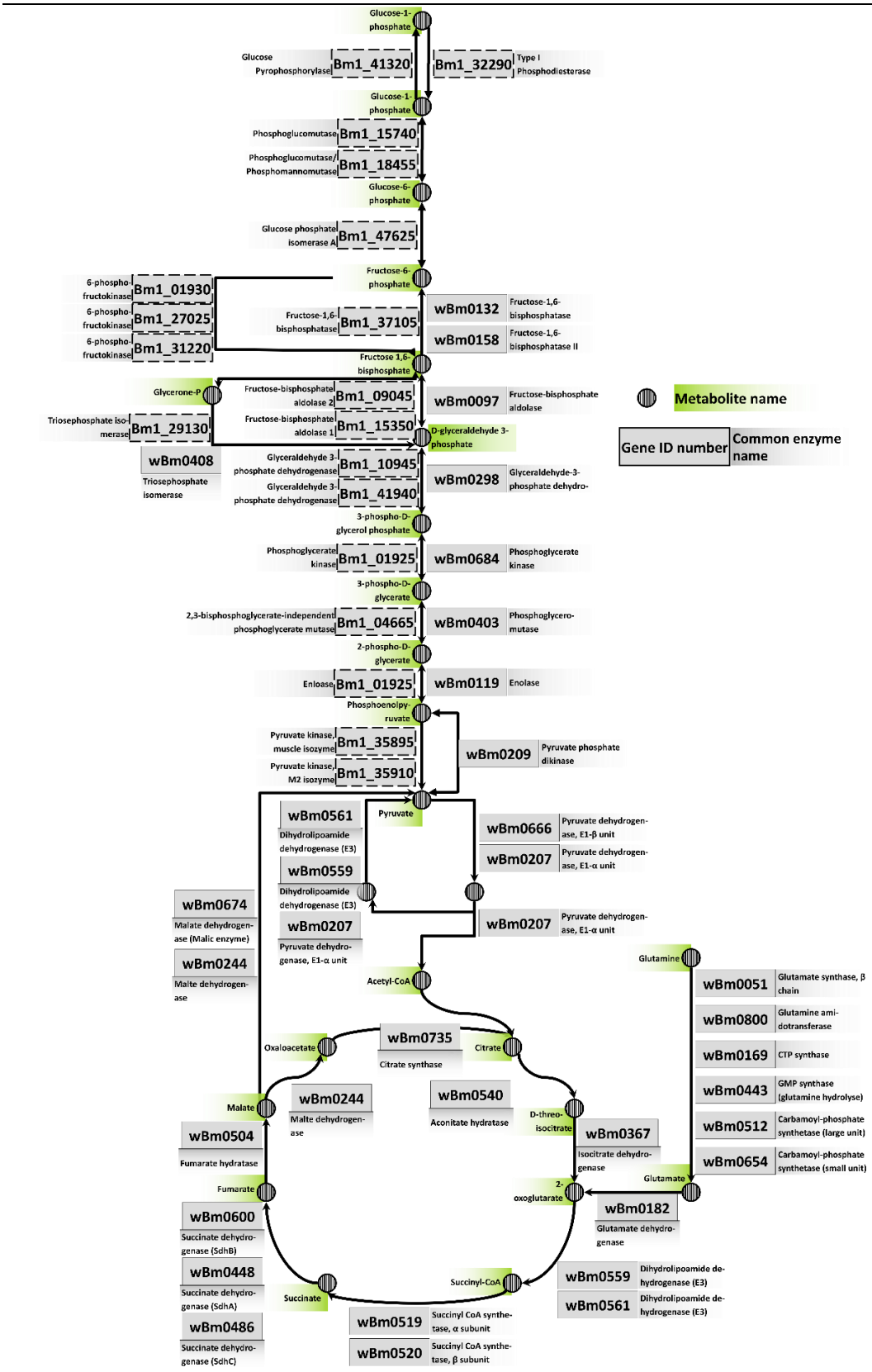
Differential expression analysis of *Wolbachia* of *B. malayi* L3-L4 development using *edgeR*
https://figshare.com/s/f9a1a9dca41438fd8401

A 2-7: *Excel table containing 6 sheets outlining the raw results from* EdgeR*, when applied to the* Wolbachia *endosymbiont during* B. malayi *L3-L4 development*

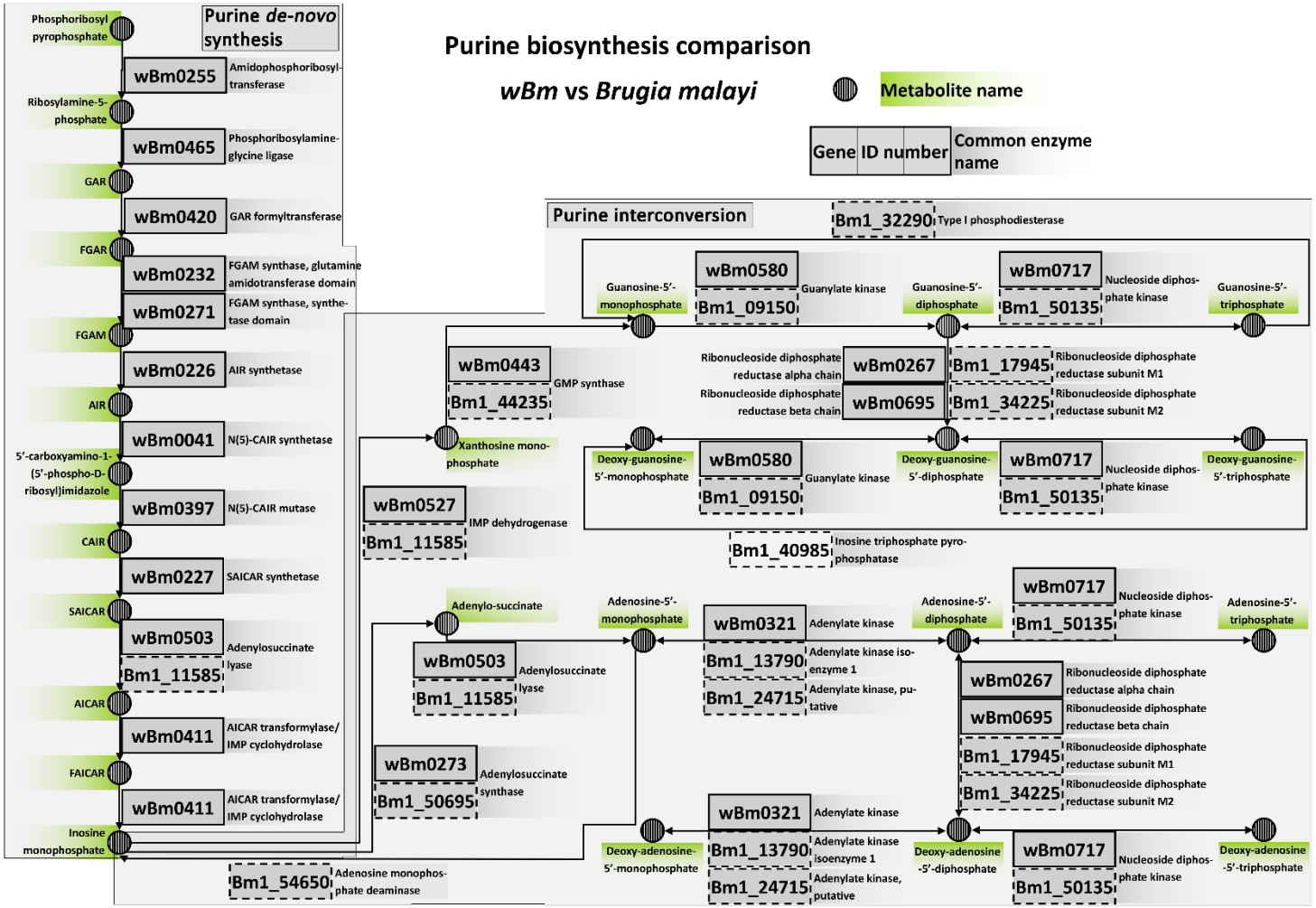
Gene Ontology enrichment tables of *B. malayi* L3-L4 development, consensus input
https://figshare.com/s/278068f8c90cbc37c960

A 2-8: *Excel table containing 6 sheets outlining the full results of Gene Ontology term enrichment of the L3-L4 developmental transcriptome of* B. malayi*.*

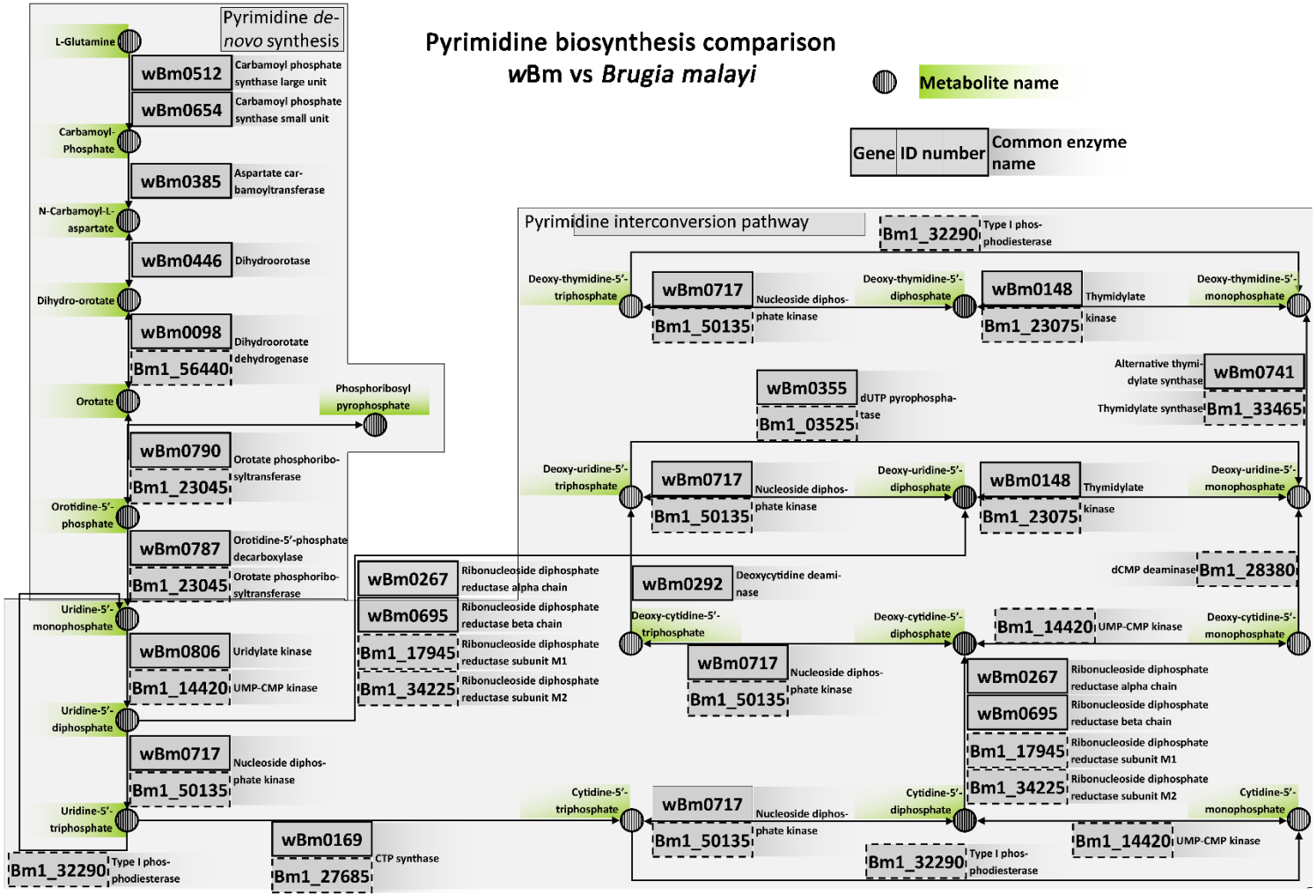
Appendix 3 Reconstructed pathways of interest in *w*Bm and *Brugia malayi*



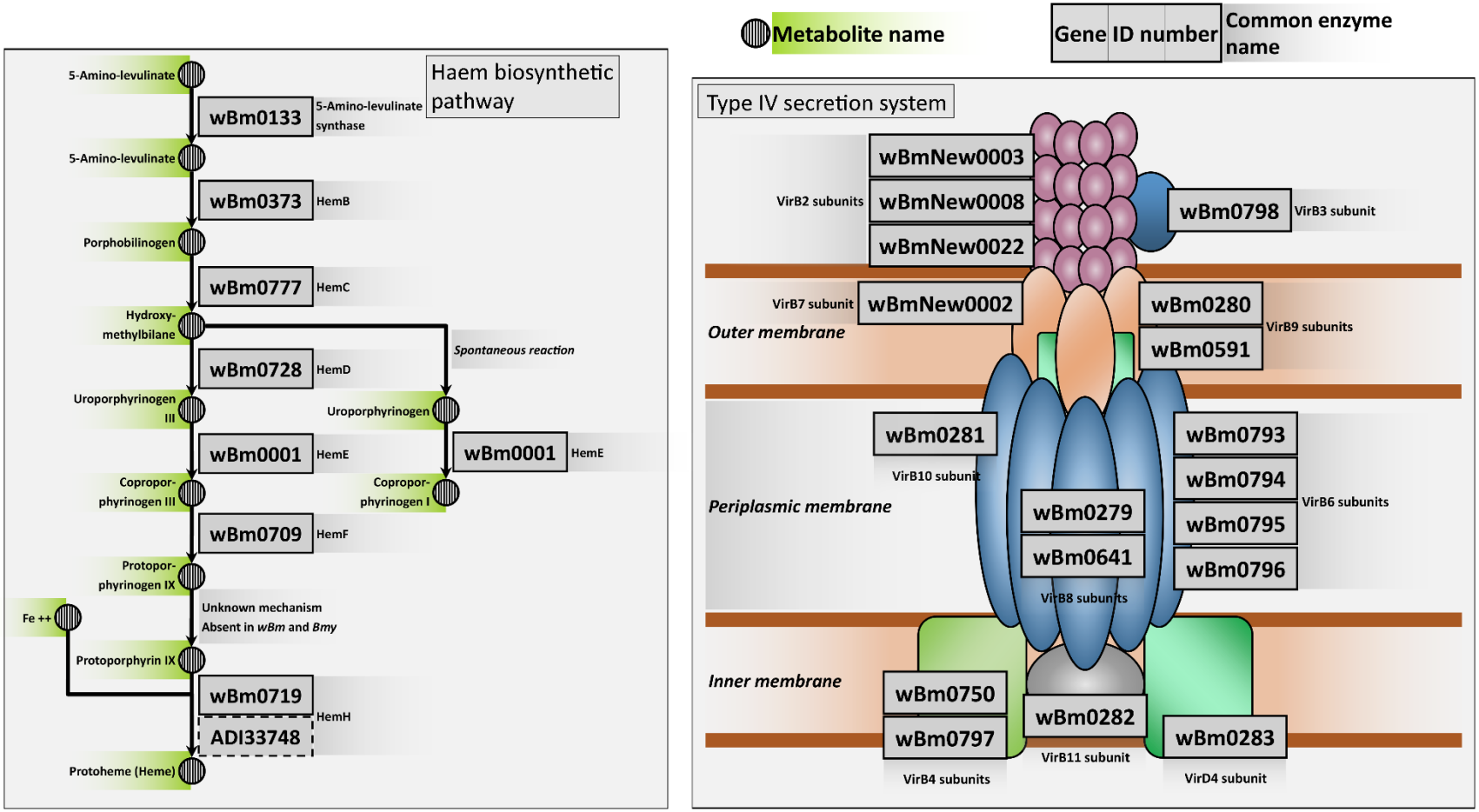
A 3-1: *Representation of the carbon metabolic pathways of* wBm*, with addition of the* B. malayi *glycolysis pathway (boxes with dotted outline) to illustrate differences between the two systems.*



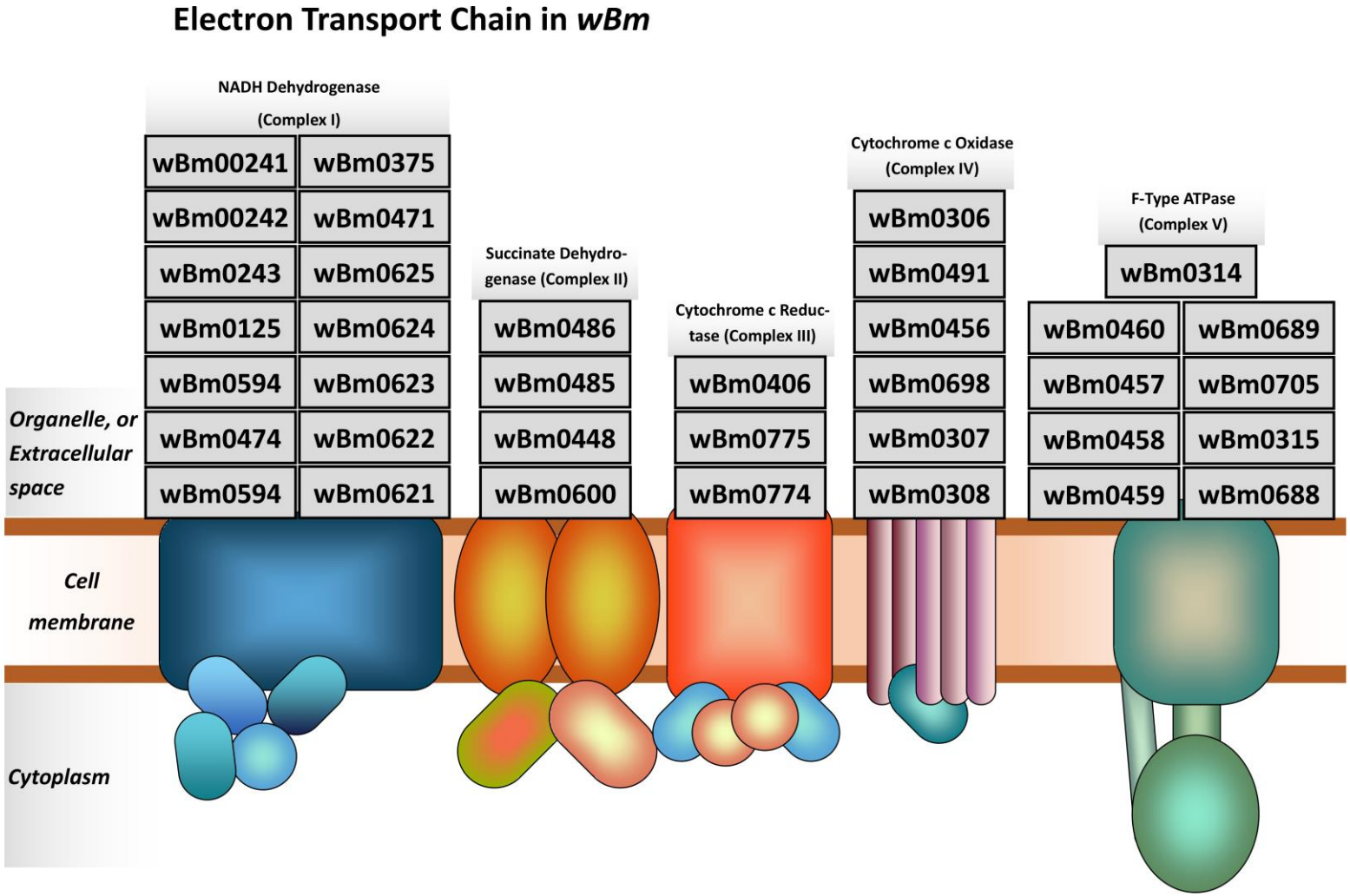
A 3-2: *Representation of the* w*Bm and* Brugia malayi *purine biosynthetic and interconversion pathways. Genes from* B. malayi *have a dotted outline, while genes from* w*Bm are solid.*



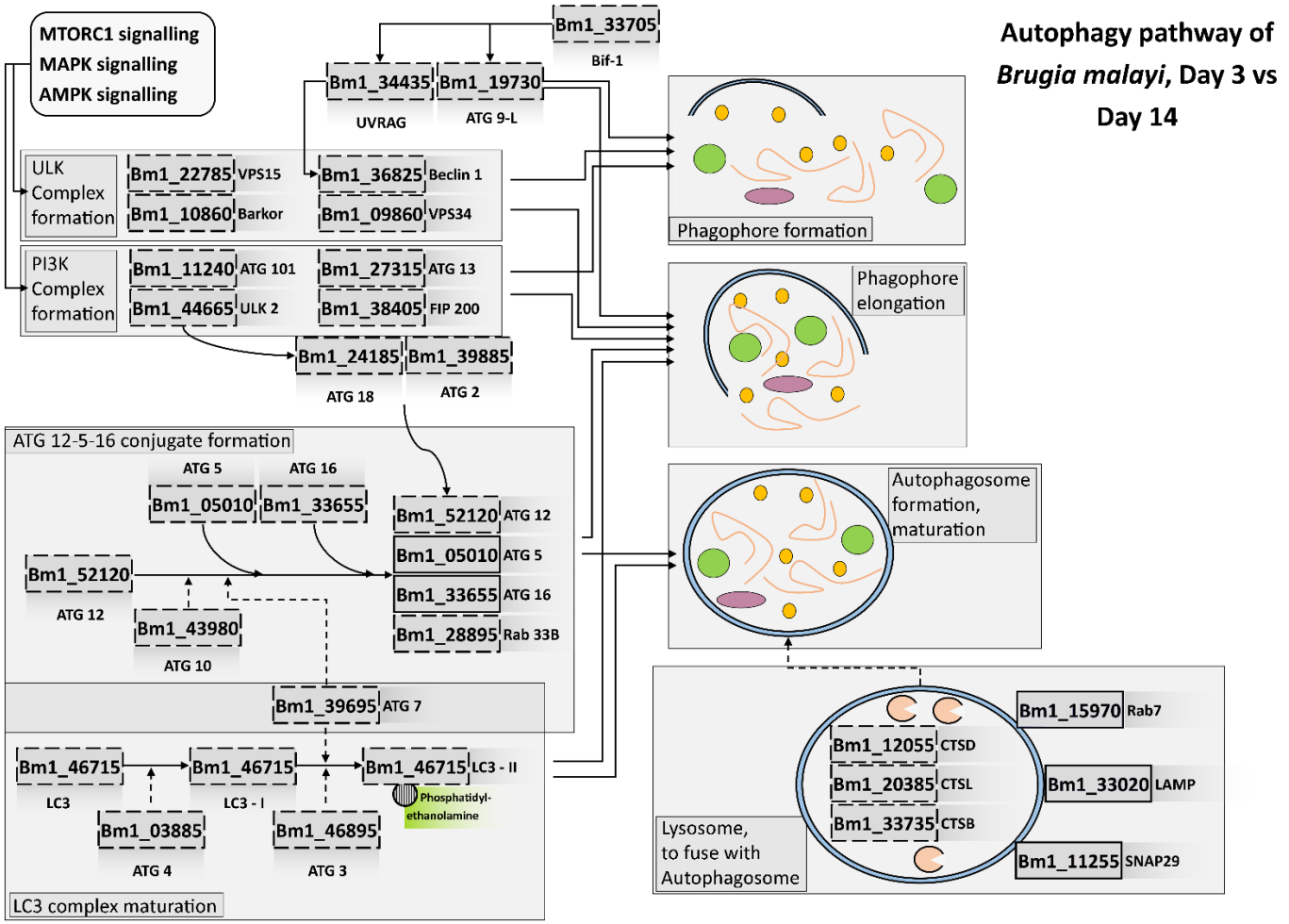
A 3-3: *Representation of the* wBm *and* Brugia malayi *pyrimidine biosynthetic and interconversion pathways. Genes from* B. malayi *have a dotted outline, while genes from* wBm *are solid.*



A 3-4: *(Left) Representation of the* w*Bm and* Brugia malayi *haem biosynthetic pathway. (Right) Representation of the* w*Bm Type IV secretion system components, with newly identified genes included.*



A 3-5: *Representation of the wBm oxidative phosphorylation system.*



A 3-6: *Representation of the* Brugia malayi *autophagy signalling pathway.*

Appendix 4 Differential expression results of the microfilariae transcriptome with and without *w*Bm depletion

In the interests of space and paper, data for this appendix section has been uploaded to the online repository, https://figshare.com/#. Digital Object Identifier (DOI) links for each item have been generated and provided for browsing.

Differential expression analysis of *B. malayi* microfilariae with and without depletion of *Wolbachia* using *edgeR*
https://figshare.com/s/67be6628e65490407a91

A 4-1: *Excel table containing 6 sheets outlining the raw results from edgeR, when applied to the transcriptome of* B. malayi *microfilariae with and without tetracycline treatment to remove the* Wolbachia *endosymbiont.*

Gene Ontology enrichment table of *B. malayi* microfilariae with and without depletion of *Wolbachia*
https://figshare.com/s/6c6b502bcb51f6ccac59

A 4-2: *Excel table containing 1 sheets outlining the full results of Gene Ontology term enrichment of the transcriptome of* B. malayi *microfilariae with and without tetracycline treatment to remove the* Wolbachia *endosymbiont.*

References

1. World Health Organization (WHO). *Weekly epidemiological record Global programme to eliminate lymphatic filariasis: progress report, 2017*. vol. 93 (2018).

2. World Health Organization (WHO). *Weekly epidemiological record Progress report on the elimination of human onchocerciasis, 2017-2018*. vol. 5 http://www.who.int/wer/2009/wer8440.pdf?ua=1 (2018).

3. James, S. L. *et al.* Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **392**, 1789–1858 (2018).

4. Taylor, M. J., Hoerauf, A. & Bockarie, M. Lymphatic filariasis and onchocerciasis. *Lancet* **376**, 1175–1185 (2010).

5. Klei, T. R. & Rajan, T. V. *World Class Parasites: Volume 5 The Filaria*. vol. 5 (Kluwer Academic Publishers, 2002).

6. World Health Organization (WHO). Lymphatic filariasis Status of Mass Drug Administration: 2018. http://apps.who.int/neglected_diseases/ntddata/lf/lf.html (2018).

7. Institute for Health Metrics and Evaluation (IHME). *Findingings from the Global Burden of Disease Study 2017*. *The Lancet* (2018).

8. Dreyer, G., Norões, J., Figueredo-Silva, J. & Piessens, W. F. Pathogenesis of lymphatic disease in bancroftian filariasis: A Clinical Perspective. *Parasitol. Today* **16**, 544–548 (2000).

9. World Health Organization (WHO). World : Distribution of lymphatic filariasis and status of preventive chemotherapy in endemic countries, 2016. http://gamapserver.who.int/mapLibrary/Files/Maps/LF_2016.png (2017).

10. Intakhan, N. & Jariyapan, N. Exsheathment and midgut invasion of nocturnally subperiodic Brugia malayi microfilariae in a refractory vector , Aedes aegypti (Thailand strain). *Parasitol. Res.* 4141–4149 (2014) doi:10.1007/s00436-014-4086-3.

11. Jariyapan, N. *et al.* Peritrophic matrix formation and Brugia malayi microfilaria invasion of the midgut of a susceptible vector, Ochlerotatus togoi (Diptera: Culicidae). *Parasitol. Res.* **112**, 2431–2440 (2013).

12. Aguldeo-Silva, F. & Spielman, A. Penetration of Mosquito Midgut Wall by Sheathed Microfilariae. *J. Invertebr. Pathol.* **45**, 117–119 (1985).

13. Erickson, S. M. *et al.* Mosquito infection responses to developing filarial worms. *PLoS Negl. Trop. Dis.* **3**, (2009).

14. Southwick, F. *Infectious Diseases: A Clinical Short Course*. *Infectious Diseases A Clinical Short Course* (McGraw-Hill, 2007). doi:10.1036/0071477225.

15. McGarry, H. F., Egerton, G. L. & Taylor, M. J. Population dynamics of Wolbachia bacterial endosymbionts in Brugia malayi. *Mol. Biochem. Parasitol.* **135**, 57–67 (2004).

16. Centers for Disease Control and Prevention. Biology - Life Cycle of Brugia malayi. https://www.cdc.gov/parasites/lymphaticfilariasis/biology_b_malayi.html (2018).

17. Burnham, G. Onchocerciasis. *Lancet* **351**, 1341–1346 (1998).

18. Ogunrinade, A., Boakye, D., Merriweather, A. & Unnasch, T. R. Distribution of the Blinding and Nonblinding strains of Onchocerca volvulus in Nigeria. *J. Infect. Dis.* **179**, 1577–1579 (1999).

19. World Health Organization (WHO). World : Distribution of onchocerciasis and status of preventive chemotherapy in endemic countries, 2017. http://gamapserver.who.int/mapLibrary/Files/Maps/Onchocerciasis_2017.png (2018).

20. Centers for Disease Control and Prevention. Biology - Life Cycle of Onchocerca volvulus. https://www.cdc.gov/parasites/onchocerciasis/biology.html (2013).

21. World Health Organization (WHO). Global Programme to Eliminate Lymphatic Filariasis: Progress Report 2000-2009 and Strategic Plan 2010-2020. *World Heal. Organ.* 1–93 (2010).

22. Niessen, L. W. & Taylor, M. Eliminating lymphatic filariasis - Is it worth it? *Clin. Infect. Dis.* 1–2 (2019) doi:10.1093/cid/ciz674.

23. Slatko, B. E., Taylor, M. J. & Foster, J. M. The Wolbachia endosymbiont as an anti-filarial nematode target. *Symbiosis* **51**, 55–65 (2010).

24. Taylor, M. J. *et al.* Onchocerciasis Control: Vision for the Future from a Ghanian perspective. *Parasit. Vectors* **2**, 7 (2009).

25. Osei-Atweneboana, M. Y. *et al.* Phenotypic Evidence of Emerging Ivermectin Resistance in Onchocerca volvulus. *PLoS Negl. Trop. Dis.* **5**, e998 (2011).

26. Awadzi, K. *et al.* An investigation of persistent microfilaridermias despite multiple treatments with ivermectin, in two onchocerciasis-endemic foci in Ghana. *Ann. Trop. Med. Parasitol.* **98**, 231–249 (2004).

27. Twum-Danso, N. A. Loa loa encephalopathy temporally related to ivermectin administration reported from onchocerciasis mass treatment programs from 1989 to 2001: implications for the future. *Filaria J.* **2 Suppl 1**, S7 (2003).

28. Senyonjo, L. *et al.* Factors Associated with Ivermectin Non-Compliance and Its Potential Role in Sustaining Onchocerca volvulus Transmission in the West Region of Cameroon. *PLoS Negl. Trop. Dis.* **10**, 1–16 (2016).

29. Kozek, W. J. & Marroquin, H. F. Intracytoplasmic Bacteria in Onchocerca volvulus. *Am. J. Trop. Med. Hyg.* **26**, 663–678 (1977).

30. Taylor, M. J. & Hoerauf, A. Wolbachia bacteria of filarial nematodes. *Parasitol. Today* **15**, 437–442 (1999).

31. Werren, J. H. Biology of Wolbachia. *Annu. Rev. Entomol.* **124**, 587–609 (1997).

32. Werren, J. H., Baldo, L. & Clark, M. E. Wolbachia: Master manipulators of invertebrate biology. *Nat. Rev. Microbiol.* **6**, 741–751 (2008).

33. Haegeman, A. *et al.* An endosymbiotic bacterium in a plant-parasitic nematode: Member of a new Wolbachia supergroup. *Int. J. Parasitol.* **39**, 1045–1054 (2009).

34. Zug, R. & Hammerstein, P. Bad guys turned nice? A critical assessment of Wolbachia mutualisms in arthropod hosts. *Biol. Rev. Camb. Philos. Soc.* **90**, 89–111 (2015).

35. Lefoulon, E. *et al.* Breakdown of coevolution between symbiotic bacteria *Wolbachia* and their filarial hosts. *PeerJ* **4**, e1840 (2016).

36. Taylor, M. J., Bordenstein, S. R. & Slatko, B. Microbe Profile : Wolbachia : a sex selector , a viral protector and a target to treat filarial nematodes. *Microbiologyopen* 1345–1347 (2019) doi:10.1099/mic.0.000724.

37. Baldo, L. & Werren, J. H. Revisiting Wolbachia Supergroup Typing Based on WSP: Spurious Lineages and Discordance with MLST. *Curr. Microbiol.* **55**, 81–87 (2007).

38. Taylor, M. J., Bandi, C. & Hoerauf, A. *Wolbachia bacterial endosymbionts of filarial nematodes*. *Advances in Parasitology* vol. 60 (Elsevier Masson SAS, 2005).

39. Zhou, W., Rousset, F. & O'Neill, S. Phylogeny and PCR-based classification of Wolbachia strains using wsp gene sequences. *Proc. R. Soc. B Biol. Sci.* **265**, 509–515 (1998).

40. Ramírez-Puebla, S. T. *et al.* Species in Wolbachia? Proposal for the designation of 'Candidatus Wolbachia bourtzisii', 'Candidatus Wolbachia onchocercicola', 'Candidatus Wolbachia blaxteri', 'Candidatus Wolbachia brugii', 'Candidatus Wolbachia taylori', 'Candidatus Wolbachia collembol. *Syst. Appl. Microbiol.* **38**, 390–399 (2015).

41. Chung, M., Munro, J. B., Tettelin, H. & Dunning Hotopp, J. C. Using Core Genome Alignments To Assign Bacterial Species. *mSystems* **3**, 1–21 (2018).

42. Hertig, M. & Wolbach, S. B. Studies on Rickettsia-Like Micro-Organisms in Insects. *J. Med. Res.* **44**, 329-374.7 (1924).

43. Hertig, M. The Rickettsia, Wolbachia pipientis (gen. et sp.n.) and Associated Inclusions of the Mosquito, Culex pipiens. *Parasitology* **28**, 453–486 (1936).

44. Hilgenboecker, K., Hammerstein, P., Schlattmann, P., Telschow, A. & Werren, J. H. How many species are infected with Wolbachia? - a statistical analysis of current data. *FEMS Microbiol. Lett.* **281**, 215–220 (2008).

45. Sazama, E. J., Ouellette, S. P. & Wesner, J. S. Bacterial Endosymbionts Are Common Among, but not Necessarily Within, Insect Species. *Environ. Entomol.* **48**, 127–133 (2019).

46. Brownlie, J. C. *et al.* Evidence for metabolic provisioning by a common invertebrate endosymbiont, wolbachia pipientis, during periods of nutritional stress. *PLoS Pathog.* **5**, (2009).

47. Gill, A. C., Darby, A. C. & Makepeace, B. L. Iron Necessity: The Secret of Wolbachia's Success? *PLoS Negl. Trop. Dis.* **8**, (2014).

48. Hedges, L. M., Brownlie, J. C., O'Neill, S. L. & Johnson, K. N. Wolbachia and Virus Protection in Insects. *Science (80-.).* **322**, 702–702 (2008).

49. Martinez, J. *et al.* Symbionts Commonly Provide Broad Spectrum Resistance to Viruses in Insects: A Comparative Analysis of Wolbachia Strains. *PLoS Pathog.* **10**, (2014).

50. Johnson, K. N. The impact of Wolbachia on virus infection in mosquitoes. *Viruses* **7**, 5705–5717 (2015).

51. Teixeira, L., Ferreira, Á. & Ashburner, M. The bacterial symbiont Wolbachia induces

resistance to RNA viral infections in Drosophila melanogaster. *PLoS Biol.* **6**, 2753–2763 (2008).

52. Moreira, L. A. *et al.* A Wolbachia Symbiont in Aedes aegypti Limits Infection with Dengue, Chikungunya, and Plasmodium. *Cell* **139**, 1268–1278 (2009).

53. Frentiu, F. D., Robinson, J., Young, P. R., Mcgraw, E. A. & Neill, S. L. O. Wolbachia-Mediated Resistance to Dengue Virus Infection and Death at the Wolbachia-Mediated Resistance to Dengue Virus Infection and Death at the Cellular Level. (2010) doi:10.1371/journal.pone.0013398.

54. Nikoh, N. *et al.* Evolutionary origin of insect-Wolbachia nutritional mutualism. *Proc. Natl. Acad. Sci.* **111**, 10257–10262 (2014).

55. Hosokawa, T., Koga, R., Kikuchi, Y., Meng, X.-Y. & Fukatsu, T. Wolbachia as a bacteriocyte-associated nutritional mutualist. *Proc. Natl. Acad. Sci.* **107**, 769–774 (2010).

56. Mclaren, D. J., Worms, M. J., Laurence, B. R. & Simpson, M. G. Micro-organisms in filarial larvae (Nematoda). *Trans. R. Soc. Trop. Med. Hyg.* **69**, 509–514 (1975).

57. Sironi, M. *et al.* Molecular evidence for a close relative of the arthropod endosymbiont Wolbachia in a filarial worm. *Mol. Biochem. Parasitol.* **74**, 223–227 (1995).

58. Taylor, M. J., Bilo, K., Cross, H. F., Archer, J. P. & Underwood, A. P. 16S rDNA Phylogeny and Ultrastructural Characterization of Wolbachia Intracellular Bacteria of the Filarial Nematodes Brugia malayi, B. pahangi, and Wuchereria bancrofti. *Exp. Parasitol.* **91**, 356–361 (1999).

59. Bandi, C., Anderson, T. J. C., Genchi, C. & Blaxter, M. L. Phylogeny of Wolbachia in filarial nematodes. *Proc. R. Soc. London. Ser. B Biol. Sci.* **265**, 2407–2413 (1998).

60. Keiser, P. B. *et al.* Molecular identification of Wolbachia from the filarial nematode Mansonella perstans. *Mol. Biochem. Parasitol.* **160**, 123–128 (2008).

61. Kozek, W. J. & Raccurt, C. Ultrastructure of Mansonella ozzardi microfilaria, with a comparison of the South American (simuliid-transmitted) and the Caribbean (culicoid-transmitted) forms. *Tropenmed. Parasitol.* **34**, 38–53 (1983).

62. Foster, J. M. *et al.* Absence of Wolbachia endobacteria in the non-filariid nematodes Angiostrongylus cantonensis and A. costaricensis. *Parasit. Vectors* **1**, 31 (2008).

63. Foster, J. M. *et al.* Absence of Wolbachia endobacteria in the human parasitic nematode Dracunculus medinensis and two related Dracunculus species infecting wildlife. *Parasites and Vectors* **7**, 1–5 (2014).

64. Brown, A. M. V. *et al.* Comparative genomics of a plant-parasitic nematode endosymbiont suggest a role in nutritional symbiosis. *Genome Biol. Evol.* **7**, 2727–2746 (2015).

65. Brown, A. M. V. *et al.* Genomic evidence for plant-parasitic nematodes as the earliest Wolbachia hosts. *Sci. Rep.* **6**, 1–14 (2016).

66. Haegeman, A. *et al.* An endosymbiotic bacterium in a plant-parasitic nematode: Member of a new Wolbachia supergroup. *Int. J. Parasitol.* **39**, 1045–1054 (2009).

67. Taylor, M. J., Voronin, D., Johnston, K. L. & Ford, L. Wolbachia filarial interactions.

Cell. Microbiol. **15**, 520–526 (2013).

68. Gehringer, C. *et al.* Molecular evidence of wolbachia endosymbiosis in mansonella perstans in gabon, central africa. *J. Infect. Dis.* **210**, 1633–1638 (2014).

69. Grobusch, M. P., Kombila, M., Autenrieth, I., Mehlhorn, H. & Kremsner, P. G. No evidence of Wolbachia endosymbiosis with Loa loa and Mansonella perstans. *Parasitol. Res.* **90**, 405–408 (2003).

70. Centers for Disease Control and Prevention. Lymphatic Filariasis. https://www.cdc.gov/dpdx/lymphaticfilariasis/index.html (2017).

71. Fenn, K. & Blaxter, M. Quantification of Wolbachia bacteria in Brugia malayi through the nematode lifecycle. *Mol. Biochem. Parasitol.* **137**, 361–364 (2004).

72. Fischer, K., Beatty, W. L., Jiang, D., Weil, G. J. & Fischer, P. U. Tissue and stage-specific distribution of Wolbachia in Brugia malayi. *PLoS Negl. Trop. Dis.* **5**, (2011).

73. Landmann, F., Foster, J. M., Slatko, B. & Sullivan, W. Asymmetric wolbachia segregation during Early Brugia malayi embryogenesis determines its distribution in adult host tissues. *PLoS Negl. Trop. Dis.* **4**, (2010).

74. McGarry, H. F. *et al.* Evidence against _Wolbachia _symbiosis in _Loa loa_. *Filaria.J* **2**, 9 (2003).

75. Bosshardt, S. C. *et al.* Prophylactic Activity of Tetracycline against Brugia pahangi Infection in Jirds (Meriones unguiculatus). *J. Parasitol.* **79**, 775 (1993).

76. Hoerauf, A. *et al.* Tetracycline therapy targets intracellular bacteria in the filarial nematode Litomosoides sigmodontis and results in filarial infertility. *J. Clin. Invest.* **103**, 11–18 (1999).

77. Bandi, C. *et al.* Effects of tetracycline on the filarial worms Brugia pahangi and Dirofilaria immitis and their bacterial endosymbionts Wolbachia. *Int. J. Parasitol.* **29**, 357–364 (1999).

78. Langworthy, N. G. *et al.* Macrofilaricidal activity of tetracycline against the filarial nematode Onchocerca ochengi: elimination of Wolbachia precedes worm death and suggests a dependent relationship. *Proc. R. Soc. B Biol. Sci.* **267**, 1063–1069 (2000).

79. Hoerauf, A. *et al.* Endosymbiotic bacteria in worms as targets for a novel chemotherapy in filariasis. *Lancet* **355**, 1242–1243 (2000).

80. Rao, R. U., Moussa, H. & Weil, G. J. Brugia malayi : effects of antibacterial agents on larval viability and development in vitro. **101**, 77–81 (2002).

81. Taylor, M. J. *et al.* Macrofilaricidal activity after doxycycline treatment of Wuchereria bancrofti: a double-blind, randomised placebo-controlled trial. *Lancet* **365**, 2116–2121 (2005).

82. Hoerauf, A. *et al.* Doxycycline as a novel strategy against bancroftian filariasis - Depletion of Wolbachia endosymbionts from Wuchereria bancrofti and stop of microfilaria production. *Med. Microbiol. Immunol.* **192**, 211–216 (2003).

83. Supat Sucharit, Somjai Viraboonchai, Narong Panavut & Chamlong Harinasuta. Studies on the effects of tetracycline on Brugia pahangi infection in Aedes togoi. *Southeast Asian J. Trop. Med. Public Health* **9**, 55–9 (1978).

84. Srivastava, K. & Misra-Bhattacharya, S. Tetracycline, a tool for transmission blocking

of Brugia malayi in Mastomys coucha. *Curr. Sci.* **85**, 588–589 (2003).

85. Arumugam, S., Pfarr, K. M. & Hoerauf, A. Infection of the intermediate mite host with Wolbachia-depleted Litomosoides sigmodontis microfilariae: Impaired L1 to L3 development and subsequent sex-ratio distortion in adult worms. *Int. J. Parasitol.* **38**, 981–987 (2008).

86. Albers, A. *et al.* Retarded Onchocerca volvulus L1 to L3 larval development in the Simulium damnosum vector after anti-wolbachial treatment of the human host. *Parasit. Vectors* 1–10 (2012).

87. Taylor, M. J., Cross, H. F. & Bilo, K. Inflammatory responses induced by the filarial nematode Brugia malayi are mediated by lipopolysaccharide-like activity from endosymbiotic Wolbachia bacteria. *J. Exp. Med.* **191**, 1429–36 (2000).

88. Turner, J. D. *et al.* Wolbachia lipoprotein stimulates innate and adaptive immunity through toll-like receptors 2 and 6 to induce disease manifestations of filariasis. *J. Biol. Chem.* **284**, 22364–22378 (2009).

89. Brattig, N. W. *et al.* The Major Surface Protein of Wolbachia Endosymbionts in Filarial Nematodes Elicits Immune Responses through TLR2 and TLR4. *J. Immunol.* **173**, 437–445 (2004).

90. Bouchery, T., Lefoulon, E., Karadjian, G., Nieguitsila, A. & Martin, C. The symbiotic role of Wolbachia in Onchocercidae and its impact on filariasis. *Clin. Microbiol. Infect.* **19**, 131–140 (2013).

91. Gillette-Ferguson, I. *et al.* Wolbachia-Induced Neutrophil Activation in a Mouse Model of Ocular Onchocerciasis (River Blindness). *Infect. Immun.* **72**, 5687–5692 (2004).

92. Andre, A. v. S. *et al.* The Role of Endosymbiotic Wolbachia Bacteria in the Pathogenesis of River Blindness. *Science (80-.).* **295**, 1892–1895 (2002).

93. Brattig, N. W., Büttner, D. W. & Hoerauf, A. Neutrophil accumulation around Onchocerca worms and chemotaxis of neutrophils are dependent on Wolbachia endobacteria. *Microbes Infect.* **3**, 439–446 (2001).

94. Nfon, C. K. *et al.* Eosinophils contribute to killing of adult Onchocerca ochengi within onchocercomata following elimination of Wolbachia. *Microbes Infect.* **8**, 2698–2705 (2006).

95. Turner, J. D. *et al.* Wolbachia endosymbiotic bacteria of Brugia malayi mediate macrophage tolerance to TLR- and CD40-specific stimuli in a MyD88/TLR2-dependent manner. *J Immunol* **177**, 1240–1249 (2006).

96. Taylor, M. J. *et al.* Macrofilaricidal activity after doxycycline treatment of Wuchereria bancrofti: a double-blind, randomised placebo-controlled trial. *Lancet* **365**, 2116–2121 (2005).

97. Debrah, A. Y. *et al.* Assessment of microfilarial loads in the skin of onchocerciasis patients after treatment with different regimens of doxycycline plus ivermectin. *Filaria J.* **5**, 1 (2006).

98. Debrah, A. Y. *et al.* Macrofilaricidal effect of 4 weeks of treatment with doxycycline on Wuchereria bancrofti. *Trop. Med. Int. Heal.* **12**, 1433–1441 (2007).

99. Mand, S. *et al.* Doxycycline improves filarial lymphedema independent of active

filarial infection: A randomized controlled trial. *Clin. Infect. Dis.* **55**, 621–630 (2012).

100. Turner, J. D. *et al.* A Randomized , Double-Blind Clinical Trial of a 3-Week Course of Doxycycline plus Albendazole and Ivermectin for the Treatment of Wuchereria bancrofti Infection. *Clin. Infect. Dis.* **42**, (2006).

101. Supali, T. *et al.* Doxycycline Treatment of *Brugia malayi* –Infected Persons Reduces Microfilaremia and Adverse Reactions after Diethylcarbamazine and Albendazole Treatment. *Clin. Infect. Dis.* **46**, 1385–1393 (2008).

102. Taylor, M. J., Hoerauf, A., Townson, S., Slatko, B. E. & Ward, S. A. Anti- Wolbachia drug discovery and development: safe macrofilaricides for onchocerciasis and lymphatic filariasis. *Parasitology* **141**, 119–127 (2013).

103. Specht, S. *et al.* Efficacy of 2- and 4-week rifampicin treatment on the Wolbachia of Onchocerca volvulus. *Parasitol. Res.* **103**, 1303–1309 (2008).

104. Debrah, A. Y. *et al.* Macrofilaricidal Activity in Wuchereria bancrofti after 2 Weeks Treatment with a Combination of Rifampicin plus Doxycycline. *J. Parasitol. Res.* **2011**, 1–9 (2011).

105. Aljayyoussi, G. *et al.* Short-course, high-dose rifampicin achieves Wolbachia depletion predictive of curative outcomes in preclinical models of lymphatic filariasis and onchocerciasis. *Sci. Rep.* **7**, 1–11 (2017).

106. Turner, J. D. *et al.* Albendazole and antibiotics synergize to deliver short-course anti-Wolbachia curative treatments in preclinical models of filariasis. *Proc. Natl. Acad. Sci.* **114**, E9712–E9721 (2017).

107. Johnston, K. L. *et al.* Identification and prioritization of novel anti-Wolbachia chemotypes from screening a 10,000-compound diversity library. *Sci. Adv.* **3**, 1–11 (2017).

108. Clare, R. H. *et al.* Industrial scale high-throughput screening delivers multiple fast acting macrofilaricide. *Nat. Commun.* **10**, (2019).

109. Hong, W. D. *et al.* AWZ1066S, a highly specific anti- Wolbachia drug candidate for a short-course treatment of filariasis. *Proc. Natl. Acad. Sci.* **116**, 1414–1419 (2019).

110. Taylor, M. J. *et al.* Preclinical development of an oral anti-Wolbachia macrolide drug for the treatment of lymphatic filariasis and onchocerciasis. *Sci. Transl. Med.* **11**, 1–11 (2019).

111. Foster, J. *et al.* The Wolbachia genome of Brugia malayi: Endosymbiont evolution within a human pathogenic nematode. *PLoS Biol.* **3**, 0599–0614 (2005).

112. Ghedin, E. *et al.* Draft Genome of the Filarial Nematode Parasite Brugia malayi. *Science (80-.).* **317**, 1756–1760 (2007).

113. Shendure, J. *et al.* DNA sequencing at 40: past, present and future. *Nature* **550**, 345–353 (2017).

114. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74**, 5463–5467 (1977).

115. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).

116. Slatko, B. E., Gardner, A. F. & Ausubel, F. M. Overview of Next-Generation

Sequencing Technologies. *Curr. Protoc. Mol. Biol.* **122**, e59 (2018).

117. Baker, M. De novo genome assembly: What every biologist should know. *Nat. Methods* **9**, 333–337 (2012).

118. Sinha, A., Li, Z., Sun, L. & Carlow, C. K. S. Complete Genome Sequence of the Wolbachia wAlbB Endosymbiont of Aedes albopictus. *Genome Biol. Evol.* **11**, 706–720 (2019).

119. Masonbrink, R. *et al.* The genome of the soybean cyst nematode (Heterodera glycines) reveals complex patterns of duplications involved in the evolution of parasitism genes. 1–14 (2019).

120. Tatusova, T. *et al.* NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **44**, 6614–6624 (2016).

121. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–10 (1990).

122. Jones, P. *et al.* InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).

123. Bumgarner, R. Overview of DNA Microarrays: Types, Applications, and Their Future. in *Current Protocols in Molecular Biology* (John Wiley & Sons, Inc., 2013). doi:10.1002/0471142727.mb2201s101.

124. Garber, M., Grabherr, M. G., Guttman, M. & Trapnell, C. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat. Methods* **8**, 469–477 (2011).

125. Zhao, S., Fung-Leung, W.-P., Bittner, A., Ngo, K. & Liu, X. Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. *PLoS One* **9**, e78644 (2014).

126. Oshlack, A., Robinson, M. D. & Young, M. D. From RNA-seq reads to differential expression results. *Genome Biol.* **11**, 1–10 (2010).

127. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).

128. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).

129. Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31**, 46–53 (2012).

130. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).

131. Consortium, T. G. O. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).

132. The Gene Ontology Consortium. The Gene Ontology Resource : 20 years and still GOing strong. *Nucleic Acids Res.* **47**, 330–338 (2019).

133. Nam, D. & Kim, S.-Y. Gene-set approach for expression pattern analysis. *Brief. Bioinform.* **9**, 189–197 (2008).

134. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545–15550 (2005).

135. Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D. & Woolf, P. J. GAGE: Generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics* **10**, 1–17 (2009).

136. Wu, M. *et al.* Phylogenomics of the reproductive parasite Wolbachia pipientis wMel: A streamlined genome overrun by mobile genetic elements. *PLoS Biol.* **2**, 327–341 (2004).

137. Kitts, P. A. *et al.* Assembly: a resource for assembled genomes at NCBI. *Nucleic Acids Res.* **44**, D73–D80 (2016).

138. Lefoulon, E. *et al.* Large Enriched Fragment Targeted Sequencing (LEFT-SEQ) Applied to Capture of Wolbachia Genomes. *Sci. Rep.* 1–10 (2019) doi:10.1038/s41598-019-42454-w.

139. Blattner, F. R. *et al.* The Complete Genome Sequence of Escherichia coli K-12. *Science (80-.).* **277**, 1453–1462 (1997).

140. Darby, A. C., Cho, N. H., Fuxelius, H. H., Westberg, J. & Andersson, S. G. E. Intracellular pathogens go extreme: genome evolution in the Rickettsiales. *Trends Genet.* **23**, 511–520 (2007).

141. Andersson, S. G. . & Kurland, C. G. Reductive evolution of resident genomes. *Trends Microbiol.* **6**, 263–268 (1998).

142. Darby, A. C. *et al.* Analysis of gene expression from the Wolbachia genome of a filarial nematode supports both metabolic and defensive roles within the symbiosis. *Genome Res.* **22**, 2467–2477 (2012).

143. Klasson, L. *et al.* Genome Evolution of Wolbachia Strain wPip from the Culex pipiens Group. *Mol. Biol. Evol.* **25**, 1877–1887 (2008).

144. Bordenstein, S. R., Marshall, M. L., Fry, A. J., Kim, U. & Wernegreen, J. J. The tripartite associations between bacteriophage, Wolbachia, and arthropods. *PLoS Pathog.* **2**, 384–393 (2006).

145. Tanaka, K., Furukawa, S., Nikoh, N., Sasaki, T. & Fukatsu, T. Complete WO phage sequences reveal their dynamic evolutionary trajectories and putative functional elements required for integration into the Wolbachia genome. *Appl. Environ. Microbiol.* **75**, 5676–5686 (2009).

146. Bordenstein, S. R. & Bordenstein, S. R. Eukaryotic association module in phage WO genomes from Wolbachia. *Nat. Commun.* **7**, 1–10 (2016).

147. Wright, J. D., Sjöstrand, F. S., Portaro, J. K. & Barr, A. R. The ultrastructure of the rickettsia-like microorganism Wolbachia pipientis and associated virus-like bodies in the mosquito Culex pipiens. *J. Ultrastruct. Res.* **63**, 79–85 (1978).

148. Beckmann, J. F., Ronau, J. A. & Hochstrasser, M. A Wolbachia deubiquitylating enzyme induces cytoplasmic incompatibility. *Nat. Microbiol.* **2**, 1–7 (2017).

149. LePage, D. P. *et al.* Prophage WO genes recapitulate and enhance Wolbachia-induced cytoplasmic incompatibility. *Nature* **543**, 243–247 (2017).

150. Lindsey, A. R. I. *et al.* Evolutionary Genetics of Cytoplasmic Incompatibility Genes cifA and cifB in Prophage WO of Wolbachia. *Genome Biol. Evol.* **10**, 434–451 (2018).

151. Landmann, F. *et al.* Both asymmetric mitotic segregation and cell-to-cell invasion are required for stable germline transmission of Wolbachia in filarial nematodes. *Biol. Open* **1**, 536–547 (2012).

152. Melnikow, E. *et al.* A Potential Role for the Interaction of Wolbachia Surface Proteins with the Brugia malayi Glycolytic Enzymes and Cytoskeleton in Maintenance of Endosymbiosis. *PLoS Negl. Trop. Dis.* **7**, (2013).

153. Gouin, E. *et al.* The RickA protein of Rickettsia conorii activates the Arp2/3 complex. *Nature* **427**, 457–461 (2004).

154. Jeng, R. L. *et al.* A Rickettsia WASP-like protein activates the Arp2/3 complex and mediates actin-based motility. *Cell. Microbiol.* **6**, 761–769 (2004).

155. Godel, C. *et al.* The genome of the heartworm, Dirofilaria immitis, reveals drug and vaccine targets. *FASEB J.* **26**, 4650–4661 (2012).

156. Raverdy, S., Foster, J. M., Roopenian, E. & Carlow, C. K. S. The Wolbachia endosymbiont of Brugia malayi has an active pyruvate phosphate dikinase. *Mol. Biochem. Parasitol.* **160**, 163–166 (2008).

157. Wu, C., Dunaway-Mariano, D. & Mariano, P. S. Design, synthesis, and evaluation of inhibitors of pyruvate phosphate dikinase. *J. Org. Chem.* **78**, 1910–1922 (2013).

158. Palayam, M., Lakshminarayanan, K., Radhakrishnan, M. & Krishnaswamy, G. Preliminary analysis to target pyruvate phosphate dikinase from wolbachia endosymbiont of Brugia malayi for designing anti-filarial agents. *Interdiscip. Sci. Comput. Life Sci.* **4**, 74–82 (2012).

159. Voronin, D. *et al.* Wolbachia lipoproteins: abundance, localisation and serology of Wolbachia peptidoglycan associated lipoprotein and the Type IV Secretion System component, VirB6 from Brugia malayi and Aedes albopictus. *Parasit. Vectors* **7**, 462 (2014).

160. Ghedin, E., Wang, S., Foster, J. M. & Slatko, B. E. First sequenced genome of a parasitic nematode. *Trends Parasitol.* **20**, 151–153 (2004).

161. International Helminth Genomes Consortium. Comparative genomics of the major parasitic worms. *Nat. Genet.* **51**, 163–174 (2019).

162. Ioannidis, P. *et al.* Extensively duplicated and transcriptionally active recent lateral gene transfer from a bacterial Wolbachia endosymbiont to its host filarial nematode Brugia malayi. *BMC Genomics* **14**, 639 (2013).

163. Dunning Hotopp, J. C. *et al.* Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science (80-.).* **317**, 1753–1756 (2007).

164. McNulty, S. N. *et al.* Endosymbiont DNA in endobacteria-free filarial nematodes indicates ancient horizontal genetic transfer. *PLoS One* **5**, (2010).

165. Henkle-Dührsen, K., Eckelt, V. H. O., Wildenburg, G., Blaxter, M. & Walter, R. D. Gene structure, activity and localization of a catalase from intracellular bacteria in Onchocerca volvulus1Note: Nucleotide sequence data reported in this paper are available in the EMBL, GenBank and DDJB databases under the accession numbers X82176, AF06. *Mol. Biochem. Parasitol.* **96**, 69–81 (1998).

166. Cotton, J. A. *et al.* The genome of Onchocerca volvulus, agent of river blindness. *Nat. Microbiol.* **2**, 1–12 (2016).

167. Luck, A. N. *et al.* Heme acquisition in the parasitic filarial nematode Brugia malayi. *FASEB J.* **30**, 3501–3514 (2016).

168. Chung, M. *et al.* Multispecies Transcriptomics Data Set of Brugia malayi , Its Wolbachia Endosymbiont wBm, and Aedes aegypti across the B. malayi Life Cycle. *Microbiol. Resour. Announc.* **7**, 1–4 (2018).

169. Grote, A. *et al.* Defining Brugia malayi and Wolbachia symbiosis by stage-specific dual RNA-seq. *PLoS Negl. Trop. Dis.* **11**, 1–21 (2017).

170. Luck, A. N. *et al.* Concurrent transcriptional profiling of Dirofilaria immitis and its Wolbachia endosymbiont throughout the nematode life cycle reveals coordinated gene expression. *BMC Genomics* 1–18 (2014).

171. Luck, A. N. *et al.* Tissue-specific transcriptomics and proteomics of a filarial nematode and its Wolbachia endosymbiont. *BMC Genomics* **16**, 920 (2015).

172. Nagaraj, V. A. *et al.* Malaria Parasite-Synthesized Heme Is Essential in the Mosquito and Liver Stages and Complements Host Heme in the Blood Stages of Infection. *PLoS Pathog.* **9**, e1003522 (2013).

173. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).

174. Tran, V. D. T. *et al.* RNA Sequencing-Based Genome Reannotation of the Dermatophyte Arthroderma benhamiae and Characterization of Its Secretome and Whole Gene Expression Profile during Infection. *mSystems* **1**, e00036-16 (2016).

175. Yang, H. *et al.* Re-annotation of eight Drosophila genomes. *Life Sci. Alliance* **1**, 1–14 (2018).

176. Delcher, A. L., Harmon, D., Kasif, S., White, O. & Salzberg, S. L. Improved microbial gene identification with Glimmer. *Nucleic Acids Res.* **27**, 4636–4641 (1999).

177. Besemer, J. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* **29**, 2607–2618 (2001).

178. Delcher, A. L., Bratke, K. A., Powers, E. C. & Salzberg, S. L. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23**, 673–679 (2007).

179. Lomsadze, A. *et al.* Modeling leaderless transcription and atypical genes results in more accurate gene prediction in prokaryotes. *Genome Res.* **28**, 1079–1089 (2018).

180. O'Neill, S. L. *et al.* In vitro cultivation of Wolbachia pipientis in an Aedes albopictus cell line. *Insect Mol. Biol.* **6**, 33–39 (1997).

181. Mavingui, P. *et al.* Whole-genome sequence of Wolbachia strain wALbB, an endosymbiont of tiger mosquito vector Aedes albopictus. *J. Bacteriol.* **194**, 1840 (2012).

182. Conesa, A. & Götz, S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* **2008**, (2008).

183. Berman, H. M. *et al.* The protein data bank. *Nucleic Acids Res.* **28**, 235–242 (2000).

184. Nawrocki, E. P. *et al.* Rfam 12.0: Updates to the RNA families database. *Nucleic Acids Res.* **43**, D130–D137 (2015).

185. Consortium, G. O. The Gene Ontology (GO) project in 2006. *Nucleic Acids Res.* **34**, D322–D326 (2006).

186. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. . Predicting transmembrane protein topology with a hidden markov model: application to complete genomes11Edited by F. Cohen. *J. Mol. Biol.* **305**, 567–580 (2001).

187. Petersen, T. N., Brunak, S., Von Heijne, G. & Nielsen, H. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786 (2011).

188. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, (2013).

189. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).

190. Rutherford, K. *et al.* Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944–945 (2000).

191. Carver, T., Harris, S. R., Berriman, M., Parkhill, J. & McQuillan, J. A. Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* **28**, 464–469 (2012).

192. Kumar, N. *et al.* Efficient Enrichment of Bacterial mRNA from Host-Bacteria Total RNA Samples. *Sci. Rep.* **6**, 1–10 (2016).

193. Luck, A. N., Slatko, B. E. & Foster, J. M. Removing the needle from the haystack: Enrichment of Wolbachia endosymbiont transcripts from host nematode RNA by Cappable-seq™. *PLoS One* **12**, 1–11 (2017).

194. Christie, P. J., Whitaker, N. & González-Rivera, C. Mechanism and structure of the bacterial type IV secretion systems. *Biochim. Biophys. Acta - Mol. Cell Res.* **1843**, 1578–1591 (2014).

195. Waksman, G. & Orlova, E. V. Structural organisation of the type IV secretion systems. *Curr. Opin. Microbiol.* **17**, 24–31 (2014).

196. Natale, P., Brüser, T. & Driessen, A. J. M. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane-Distinct translocases and mechanisms. *Biochim. Biophys. Acta - Biomembr.* **1778**, 1735–1756 (2008).

197. Du Plessis, D. J. F., Nouwen, N. & Driessen, A. J. M. The Sec translocase. *Biochim. Biophys. Acta - Biomembr.* **1808**, 851–865 (2011).

198. Mori, H. & Ito, K. The Sec protein-translocation pathway. *Trends Microbiol.* **9**, 494–500 (2001).

199. Käll, L., Krogh, A. & Sonnhammer, E. L. L. A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **338**, 1027–1036 (2004).

200. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

201. Altschul, S. F. & Koonin, E. V. Iterated profile searches with PSI-BLAST - A tool for discovery in protein databases. *Trends Biochem. Sci.* **23**, 444–447 (1998).

202. Notredame, C., Higgins, D. G. & Heringa, J. T-coffee: a novel method for fast and accurate multiple sequence alignment 1 1Edited by J. Thornton. *J. Mol. Biol.* **302**, 205–217 (2000).

203. Battaile, K. P., Nguyen, T. V., Vockley, J. & Kim, J. J. P. Structures of isobutyryl-CoA dehydrogenase and enzyme-product complex: Comparison with isovaleryl- and short-chain acyl-CoA dehydrogenases. *J. Biol. Chem.* **279**, 16526–16534 (2004).

204. Jacks, T. Translational Suppression in Gene Expression in Retroviruses and Retrotransposons. in *Retroviruses* (eds. Swanstrom, R. & Vogt, P. K.) 93–124 (Springer Berlin Heidelberg, 1990).

205. Jacks, T. & Varmus, H. E. Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. *Science (80-.).* **230**, 1237 LP – 1242 (1985).

206. Harrell, L. Predominance of six different hexanucleotide recoding signals 3' of read-through stop codons. *Nucleic Acids Res.* **30**, 2011–2017 (2002).

207. Antonov, I., Coakley, A., Atkins, J. F., Baranov, P. V. & Borodovsky, M. Identification of the nature of reading frame transitions observed in prokaryotic genomes. *Nucleic Acids Res.* **41**, 6514–6530 (2013).

208. Jungreis, I. *et al.* Evidence of abundant stop codon readthrough in Drosophila and other metazoa. *Genome Res.* **21**, 2096–2113 (2011).

209. Namy, O. *et al.* Identification of stop codon readthrough genes in Saccharomyces cerevisiae. *Nucleic Acids Res.* **31**, 2289–2296 (2003).

210. Porankiewlcz, J., Wang, J. & Clarke, A. K. New insights into the ATP-dependent Clp protease: Escherichia coli and beyond. *Mol. Microbiol.* **32**, 449–458 (1999).

211. Yu, A. Y. H. & Houry, W. A. ClpP: a distinctive family of cylindrical energy-dependent serine proteases. *FEBS Lett.* **581**, 3749—3757 (2007).

212. Gottesman, S. & Clark, W P, M. M. R. The ATP-dependent Clp Protease of Escherichia. *Jbc* **265**, 7886–7893 (1990).

213. Duran, E. C., Weaver, C. L. & Lucius, A. L. Comparative Analysis of the Structure and Function of AAA+ Motors ClpA, ClpB, and Hsp104: Common Threads and Disparate Functions. *Front. Mol. Biosci.* **4**, 54 (2017).

214. Grimaud, R., Kessel, M., Beuron, F., Steven, A. C. & Maurizi, M. R. Enzymatic and structural similarities between the Escherichia coli ATP-dependent proteases, ClpXP and ClpAP. *J. Biol. Chem.* **273**, 12476—12481 (1998).

215. Flynn, J. M., Neher, S. B., Kim, Y. I., Sauer, R. T. & Baker, T. A. Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Mol. Cell* **11**, 671—683 (2003).

216. Sahl, J. W., Caporaso, J. G., Rasko, D. A. & Keim, P. The large-scale blast score ratio (LS-BSR) pipeline: a method to rapidly compare genetic content between bacterial genomes. *PeerJ* **2**, e332 (2014).

217. Doolittle, R. F. *Of URFS and ORFS: A primer on how to analyse derived amino acid sequences*. (University Science Books, 1986).

218. Chung, M., Small, S. T., Serre, D., Zimmerman, P. A. & Hotopp, J. C. D. Draft genome sequence of the Wolbachia endosymbiont of Wuchereria bancrofti wWb. *Pathog.*

Dis. 1–6 (2017) doi:10.1093/femspd/ftx115.

219. Rawe, J. O. *et al.* Low concordance of multiple variant-calling pipelines : practical implications for exome and genome sequencing. *Genome Med.* (2013) doi:10.1186/gm432.

220. Sandmann, S., Graaf, A. O. De, Karimi, M. & Reijden, B. A. Van Der. Evaluating Variant Calling Tools for Sequencing Data. *Nat. Publ. Gr.* 1–12 (2017) doi:10.1038/srep43169.

221. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

222. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin).* **6**, 80–92 (2012).

223. Robinson, J. T. *et al.* Integrative Genomics Viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).

224. Frutos, R. *et al.* Comparative Genomic Analysis of Three Strains of Ehrlichia ruminantium Reveals an Active Process of Genome Size Plasticity. **188**, 2533–2542 (2006).

225. Rancès, E., Voronin, D., Tran-Van, V. & Mavingui, P. Genetic and functional characterization of the type IV secretion system in Wolbachia. *J. Bacteriol.* **190**, 5020–5030 (2008).

226. Alvarez-Martinez, C. E. & Christie, P. J. Biological Diversity of Prokaryotic Type IV Secretion Systems. *Microbiol. Mol. Biol. Rev.* **73**, 775–808 (2009).

227. Melnikow, E. *et al.* Interaction of a Wolbachia WSP-like protein with a nuclear-encoded protein of Brugia malayi. *Int. J. Parasitol.* **41**, 1053–1061 (2011).

228. Frank, D. N. & Pace, N. R. Ribonuclease P : Unity and Diversity in a tRNA Processing Ribozyme. *Annu. Rev. Biochem.* **67**, 153–180 (1998).

229. Trotochaud, A. E. & Wassarman, K. M. A highly conserved 6S RNA structure is required for regulation of transcription. *Nat. Struct. Mol. Biol.* **12**, 313–319 (2005).

230. Barrick, J. E., Sudarsan, N., Weinberg, Z., Ruzzo, W. L. & Breaker, R. R. 6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter. *RNA* **11**, 774–784 (2005).

231. Moore, S. D. & Sauer, R. T. The tmRNA System for Translational Surveillance and Ribosome Rescue. *Annu. Rev. Biochem.* **76**, (2007).

232. Mao, C. *et al.* Variations on the tmRNA gene. *RNA Biol.* **6**, 355–361 (2009).

233. Coletta, A. *et al.* Low-complexity regions within protein sequences have position-dependent roles. *BMC Syst. Biol.* **4**, 43 (2010).

234. Weiner, A. M. & Weber, K. Natural Read-through at the UGA Termination Signal of Qβ Coat Protein Cistron. *Nat. New Biol.* **234**, 206–209 (1971).

235. Baranov, P. V., Atkins, J. F. & Yordanova, M. M. Augmented genetic decoding: Global, local and temporal alterations of decoding processes and codon meaning. *Nat. Rev. Genet.* **16**, 517–529 (2015).

236. Atkins, J. F., Loughran, G., Bhatt, P. R., Firth, A. E. & Baranov, V. Ribosomal frameshifting and transcriptional slippage : From genetic steganography and

cryptography to adventitious use. *Nucleic Acids Res.* **44**, 7007–7078 (2016).

237. Korkmaz, G., Holm, M., Wiens, T. & Sanyal, S. Comprehensive analysis of stop codon usage in bacteria and its correlation with release factor abundance. *J. Biol. Chem.* **289**, 30334–30342 (2014).

238. Tompkins, R. & Caskey, T. Release factors differing in specificity for terminator codons. *Proc. Natl. Acad. Sci.* **2**, 768–774 (1968).

239. Nilsson, M. & Rydén-Aulin, M. Glutamine is incorporated at the nonsense codons UAG and UAA in a suppressor-free Escherichia coli strain. *Biochim. Biophys. Acta - Gene Struct. Expr.* **1627**, 1–6 (2003).

240. Waters, L. S. & Storz, G. Regulatory RNAs in Bacteria. *Cell* **136**, 615–628 (2009).

241. Storz, G., Vogel, J. & Wassarman, K. M. Regulation by Small RNAs in Bacteria: Expanding Frontiers. *Mol. Cell* **43**, 880–891 (2011).

242. Goodhead, I. & Darby, A. C. Taking the pseudo out of pseudogenes. *Curr. Opin. Microbiol.* **23**, 102–109 (2015).

243. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011).

244. Joshi, N. & Fass, J. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33). (2011).

245. Barbraham Bioinformatics. FastQC: A quality control tool for high throughput sequence data. (2011).

246. Murfin, K. E. *et al.* Nematode-Bacterium Symbioses - Cooperation and Conflict Revealed in the 'Omics' Age. **223**, 85–102 (2013).

247. Wu, B. *et al.* The heme biosynthetic pathway of the obligate Wolbachia endosymbiont of Brugia malayias a potential anti-filarial drug target. *PLoS Negl. Trop. Dis.* **3**, 1–10 (2009).

248. Smith, H. L. & Rajan, T. V. Tetracycline Inhibits Development of the Infective-Stage Larvae of Filarial Nematodes in Vitro. *Exp. Parasitol.* **270**, 265–270 (2000).

249. Li, B.-W., Wang, Z., Rush, A. C., Mitreva, M. & Weil, G. J. Transcription profiling reveals stage- and function-dependent expression patterns in the filarial nematode Brugia malayi. *BMC Genomics* **13**, 184 (2012).

250. Choi, Y. J. *et al.* A deep sequencing approach to comparatively analyze the transcriptome of lifecycle stages of the filarial worm, brugia malayi. *PLoS Negl. Trop. Dis.* **5**, (2011).

251. Grote, A., Lustigman, S. & Ghedin, E. Lessons from the genomes and transcriptomes of filarial nematodes. *Mol. Biochem. Parasitol.* **215**, 23–29 (2017).

252. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

253. Bennuru, S. *et al.* Stage-specific transcriptome and proteome analyses of the filarial parasite Onchocerca volvulus and its Wolbachia endosymbiont. *MBio* **7**, 1–11 (2016).

254. Williams, C. R., Baccarella, A., Parrish, J. Z. & Kim, C. C. Empirical assessment of

analysis workflows for differential expression analysis of human samples using RNA-Seq. *BMC Bioinformatics* **18**, 1–12 (2017).

255. Ogata, H. *et al.* KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **27**, 29–34 (1999).

256. Karp, P. D. *et al.* Pathway tools version 19.0 update: Software for pathway/genome informatics and systems biology. *Brief. Bioinform.* **17**, 877–890 (2016).

257. Paley, S., O'Maille, P. E., Weaver, D. & Karp, P. D. Pathway collages: Personalized multi-pathway diagrams. *BMC Bioinformatics* **17**, 1–10 (2016).

258. Luo, W. & Brouwer, C. Pathview: An R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* **29**, 1830–1831 (2013).

259. Powell, D. R. Degust: interactive RNA-seq analysis. http://victorian-bioinformatics-consortium.github.io/degust/ (2017) doi:10.5281/zenodo.3258932.

260. Marians, K. J., Hiasa, H., Kim, D. R. & Mchenry, C. S. Role of the Core DNA Polymerase III Subunits at the Replication Fork. *J. Biol. Chem.* **273**, 2452–2457 (1998).

261. Maki, H., Maki, S. & Kornberg, A. DNA Polymerase III holoenzyme of Escherichia coli. IV. The holoenzyme is an asymmetric dimer with twin active sites. *J. Biol. Chem.* **263**, 6570–6578 (1988).

262. Timson, D. J., Singleton, M. R. & Wigley, D. B. DNA ligases in the repair and replication of DNA. *Mutat. Res.* **460**, 301–318 (2000).

263. Ohyama, H. *et al.* The role of ribonucleases in regulating global mRNA levels in the model organism Thermus thermophilus HB8. *BMC Genomics* **15**, 386 (2014).

264. Tannous, E., Kanaya, E. & Kanaya, S. Role of RNase H1 in DNA repair: removal of single ribonucleotide misincorporated into DNA in collaboration with RNase H2. *Sci. Rep.* 1–11 (2015) doi:10.1038/srep09969.

265. Wu, B. *et al.* Interdomain lateral gene transfer of an essential ferrochelatase gene in human parasitic nematodes. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 7748–53 (2013).

266. Voronin, D., Cook, D. A. N., Steven, A. & Taylor, M. J. Autophagy regulates Wolbachia populations across diverse symbiotic associations. *Proc. Natl. Acad. Sci.* **109**, E1638–E1646 (2012).

267. Sarkar, S. Regulation of autophagy by mTOR-dependent and mTOR-independent pathways: autophagy dysfunction in neurodegenerative diseases and therapeutic application of autophagy enhancers. *Biochem. Soc. Trans.* 1103–1130 (2013) doi:10.1042/BST20130134.

268. Lipinski, M. M. *et al.* A Genome-Wide siRNA Screen Reveals Multiple mTORC1 Independent Signaling Pathways Regulating Autophagy under Normal Nutritional Conditions. *Dev. Cell* 1041–1052 (2010) doi:10.1016/j.devcel.2010.05.005.

269. Raizen, D. M. *et al.* Lethargus is a Caenorhabditis elegans sleep-like state. *Nature* **451**, 569–572 (2008).

270. Page, A. P., Stepek, G., Winter, A. D. & Pertab, D. Enzymology of the nematode cuticle: A potential drug target? *Int. J. Parasitol. Drugs Drug Resist.* **4**, 133–141 (2014).

271. Serbus, L. R., Casper-Lindley, C., Landmann, F. & Sullivan, W. The Genetics and Cell Biology of *Wolbachia* -Host Interactions. *Annu. Rev. Genet.* **42**, 683–707 (2008).

272. Vander Heiden, M., Cantley, L. & Thompson, C. Understanding the Warburg effect: The metabolic Requirements of cell proliferation. *Science (80-.).* **324**, 1029–1033 (2009).

273. Pfeiffer, T. & Morley, A. An evolutionary perspective on the Crabtree effect. *Front. Mol. Biosci.* **1**, 1–6 (2014).

274. Salcedo-Sora, J. E., Caamano-Gutierrez, E., Ward, S. A. & Biagini, G. A. The proliferating cell hypothesis: A metabolic framework for Plasmodium growth and development. *Trends Parasitol.* **30**, 170–175 (2014).

275. Vadlakonda, L., Dash, A., Pasupuleti, M., Anil Kumar, K. & Reddanna, P. Did We Get Pasteur, Warburg, and Crabtree on a Right Note? *Front. Oncol.* **3**, 1–4 (2013).

276. Voronin, D. *et al.* Glucose and glycogen metabolism in brugia malayi is associated with wolbachia symbiont fitness. *PLoS One* **11**, 1–18 (2016).

277. Voronin, D. *et al.* Pyruvate produced by Brugia spp. via glycolysis is essential for maintaining the mutualistic association between the parasite and its endosymbiont, Wolbachia. *PLOS Pathog.* **15**, e1008085 (2019).

278. Zhao, H., Zhu, M., Limbo, O. & Russell, P. RNase H eliminates R-loops that disrupt DNA replication but is nonessential for efficient DSB repair. *EMBO Rep.* **19**, 1–10 (2018).

279. Landmann, F., Voronin, D., Sullivan, W. & Taylor, M. J. Anti-filarial activity of antibiotic therapy is due to extensive apoptosis after Wolbachia depletion from filarial nematodes. *PLoS Pathog.* **7**, 1–11 (2011).

280. Garneau, N. L., Wilusz, J. & Wilusz, C. J. The highways and byways of mRNA decay. *Nat. Rev. Mol. Cell Biol.* **8**, 113–126 (2007).

281. Mutafchiev, Y., Bain, O., Williams, Z., McCall, J. W. & Michalski, M. L. Intraperitoneal development of the filarial nematode Brugia malayi in the Mongolian jird (Meriones unguiculatus). *Parasitol. Res.* **113**, 1827–1835 (2014).

282. He, C. & Klionsky, D. J. Regulation Mechanisms and Signalling Pathways of Autophagy. *Annu. Rev. Genet.* 67–93 (2009) doi:10.1146/annurev-genet-102808-114910.Regulation.

283. Glick, D., Barth, S. & Macleod, K. F. Autophagy : cellular and molecular mechanisms. *J. Pathol.* **221**, 3–12 (2010).

284. Meijer, A. J., Lorin, S., Blommaart, E. F. & Codogno, P. Regulation of autophagy by amino acids and MTOR-dependent signal transduction. *Amino Acids* 2037–2063 (2015) doi:10.1007/s00726-014-1765-4.

285. Mudunuri, U., Che, A., Yi, M. & Stephens, R. M. bioDBnet : the biological database network. *Bioinformatics* **25**, 555–556 (2009).

286. Malinich, E. A., Wang, K., Mukherjee, P. K., Kolomiets, M. & Kenerley, C. M. Differential expression analysis of Trichoderma virens RNA reveals a dynamic transcriptome during colonization of Zea mays roots. *BMC Genomics* **20**, 280 (2019).

287. Rao, R. U. *et al.* Effects of Doxycycline on gene expression in Wolbachia and Brugia

malayi adult female worms in vivo. *J. Biomed. Sci.* **19**, 21 (2012).

288. Darby, A. C. *et al.* Integrated transcriptomic and proteomic analysis of the global response of wolbachia to doxycycline-induced stress. *ISME J.* **8**, 925–937 (2014).

289. Martin, J. A. & Wang, Z. Next-generation transcriptome assembly. *Nat. Rev. Genet.* **12**, 671–682 (2011).

290. Glass, E. M. & Meyer, F. The Metagenomics RAST Server: A Public Resource for the Automatic Phylogenetic and Functional Analysis of Metagenomes. *Handb. Mol. Microb. Ecol. I Metagenomics Complement. Approaches* **8**, 325–331 (2011).

291. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).

292. Nolan, T., Hands, R. E. & Bustin, S. A. Quantification of mRNA using real-time RT-PCR. *Nat. Protoc.* **1**, 1559–1582 (2006).

293. Bustin, S. A. *et al.* The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* **55**, 611–622 (2009).

294. Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).

295. Perkins, D. N., Pappin, D. J. C., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567 (1999).

296. Strübing, U., Lucius, R., Hoerauf, A. & Pfarr, K. M. Mitochondrial genes for heme-dependent respiratory chain complexes are up-regulated after depletion of Wolbachia from filarial nematodes. *Int. J. Parasitol.* **40**, 1193–1202 (2010).

297. Cotton, J. A. *et al.* The genome of Onchocerca volvulus, agent of river blindness. *Nat. Microbiol.* **2**, (2016).

298. Armstrong, S. D. *et al.* Stage-specific Proteomes from Onchocerca ochengi, Sister Species of the Human River Blindness Parasite, Uncover Adaptations to a Nodular Lifestyle. *Mol. Cell. Proteomics* **15**, 2554–75 (2016).

299. Armstrong, S. D. *et al.* Comparative analysis of the secretome from a model filarial nematode (Litomosoides sigmodontis) reveals maximal diversity in gravid female parasites. *Mol. Cell. Proteomics* **13**, 2527–44 (2014).

300. Al Ait, L., Yamak, Z. & Morgenstern, B. DIALIGN at GOBICS--multiple sequence alignment using various sources of external information. *Nucleic Acids Res.* **41**, 3–7 (2013).

301. Camlin, N. J., McLaughlin, E. A. & Holt, J. E. Motoring through: The role of kinesin superfamily proteins in female meiosis. *Hum. Reprod. Update* **23**, 409–420 (2017).

302. Marx, A., Hoenger, A. & Mandelkow, E. Structures of kinesin motor proteins. *Cell Motil. Cytoskeleton* **66**, 958–966 (2009).

303. Turner, J. *et al.* Crystal Structure of the Mitotic Spindle Kinesin Eg5 Reveals a Novel Conformation of the Neck-linker. *J. Biol. Chem.* **276**, 25496–25502 (2001).

304. Ferenz, N. P., Gable, A. & Wadsworth, P. Mitotic functions of kinesin-5. *Semin. Cell*

Dev. Biol. **21**, 255–259 (2010).

305. Stolovitzky, G. & Cecchi, G. Efficiency of DNA replication in the polymerase chain reaction. *Proc. Natl. Acad. Sci.* **93**, 12947–12952 (1996).

306. Svec, D., Tichopad, A., Novosadova, V., Pfaffl, M. W. & Kubista, M. How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments. *Biomol. Detect. Quantif.* **3**, 9–16 (2015).

307. Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT – PCR. *Nucleic Acids Res.* **29**, 16–21 (2001).

308. Keeling, P. J. & Doolittle, W. F. Alpha-tubulin from early-diverging eukaryotic lineages and the evolution of the tubulin family. *Mol. Biol. Evol.* **13**, 1297–1305 (1996).

309. Jenkins, C. *et al.* Genes for the cytoskeletal protein tubulin in the bacterial genus Prosthecobacter. *Proc. Natl. Acad. Sci.* **99**, 17049–17054 (2002).

310. Pilhofer, M., Ladinsky, M. S., McDowall, A. W., Petroni, G. & Jensen, G. J. Microtubules in Bacteria: Ancient tubulins build a five-protofilament homolog of the eukaryotic cytoskeleton. *PLoS Biol.* **9**, (2011).

311. Dagenbach, E. M. & Endow, S. A. A new kinesin tree. *J. Cell Sci.* **117**, 3–7 (2004).

312. Lawrence, C. J. *et al.* A standardized kinesin nomenclature. *J. Cell Biol.* **167**, 19–22 (2004).

313. Vale, R. D. & Milligan, R. A. The Way Things Move: Looking Under the Hood of Molecular Motor Proteins. *Science (80-.).* **288**, 88–95 (2000).

314. Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1**, 945–951 (1982).

315. Block, S. M. Kinesin motor mechanics: Binding, stepping, tracking, gating, and limping. *Biophys. J.* **92**, 2986–2995 (2007).

316. Hart, C. L. *et al.* A structural change in the kinesin motor protein that drives motility. *Nature* **402**, 778–784 (1999).

317. Downing, K. H. & Nogales, E. Tubulin and microtubule structure. *Curr. Opin. Cell Biol.* **10**, 16–22 (1998).

318. Stueland, C. S., Gorden, K. & LaPorte, D. C. The isocitrate dehydrogenase phosphorylation cycle. Identification of the primary rate-limiting step. *J. Biol. Chem.* **263**, 19475–19479 (1988).

319. Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. A Greedy Algorithm for Aligning DNA Sequences. *J. Comput. Biol.* **7**, 203–214 (2000).

320. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

321. Trapnell, C. *et al.* Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. *Nat. Biotechnol.* **28**, 511–515 (2011).

322. Ye, J. *et al.* Primer-BLAST : A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**, (2012).

323. Bain, O. & Babayan, S. Behaviour of filariae: morphological and anatomical signatures of their life style within the arthropod and vertebrate hosts. *Filaria J.* **2**, 16 (2003).

324. Chandy, A., Thakur, A. S., Singh, M. P. & Manigauha, A. A review of neglected tropical diseases: Filariasis. *Asian Pac. J. Trop. Med.* **4**, 581–586 (2011).

325. Michalski, M. L., Erickson, S. M., Bartholomay, L. C. & Christensen, B. M. Midgut barrier imparts selective resistance to filarial worm infection in Culex Pipiens Pipiens. *PLoS Negl. Trop. Dis.* **4**, (2010).

326. Christensen, B. M. & Sutherland, D. R. Brugia pahangi: Exsheathment and Midgut Penetration in Aedes aegypti. *Trans. Am. Microsc. Soc.* **103**, 423–433 (1984).

327. Christensen, B. M., Li, J., Chen, C. & Nappi, A. J. Melanization immune responses in mosquito vectors. **21**, (2005).

328. Zhao, X., Ferdig, M. T., Li, J. & Christensen, B. M. Biochemical pathway of melanotic encapsulation of Brugia malayi in the mosquito, Armigeres subalbatus. *Dev. Comp. Immunol.* **19**, 205–215 (1995).

329. Terra, W. R. The origin and functions of the insect peritrophic membrane and peritrophic gel. *Arch. Insect Biochem. Physiol.* **47**, 47–61 (2001).

330. Merzendorfer, H. & Zimoch, L. Chitin metabolism in insects : structure , function and regulation of chitin synthases and chitinases. *J. Exp. Biol.* **206**, 4393–4412 (2003).

331. Chung, M. *et al.* Targeted enrichment outperforms other enrichment techniques and enables more multi-species RNA- Seq analyses. *Sci. Rep.* 1–12 (2018) doi:10.1038/s41598-018-31420-7.

332. Perrone, J. B. & Spielman, A. Time and site of assembly of the peritrophic membrane of the mosquito Aedes aegypti. *Cell Tissue Res.* (1988).

333. Kato, N. *et al.* Evaluation of the Function of a Type I Peritrophic Matrix as a Physical Barrier for Midgut Epithelium Invasion by Mosquito-Borne Pathogens in *Aedes aegypti*. *Vector-Borne Zoonotic Dis.* **8**, 701–712 (2008).

334. Harnett, W., Meghji, M., Worms, M. J. & Parkhouse, R. M. E. Quantitative and qualitative changes in production of excretions/secretions by Litomosoides carinii during development in the jird (Meriones unguiculatus). *Parasitology* **93**, 317–331 (1986).

335. Hintz, M. *et al.* Juvenile female Litomosoides sigmodontis produce an excretory\secretory antigen (Juv-p120) highly modified with dimethylaminoethanol. *Parasitology* **117**, 265–271 (1998).

336. Wagner, U. *et al.* Characterization of the DMAE-modified juvenile excretory – secretory protein Juv-p120 of Litomosoides sigmodontis. *Mol. Biochem. Parasitol.* **176**, 80–89 (2011).

337. Beyenbach, K. W. & Wieczorek, H. The V-type H + ATPase : molecular structure and function , physiological roles and regulation. *J. Exp. Biol.* **209**, 577–589 (2006).

338. Wu, Y., Williams, S. A. & Bianco, E. Chitinase genes expressed by infective larvae of

the filarial nematodes, Acanthocheilonema viteae and Onchocerca volvulus. *Mulecular Biochem. Parasitol.* **75**, 207–219 (1996).

339. Wu, Y., Egerton, G., Underwood, A. P., Sakuda, S. & Bianco, A. E. Expression and Secretion of a Larval-specific Chitinase (Family 18 Glycosyl Hydrolase) by the Infective Stages of the Parasitic Nematode , Onchocerca volvulus *. **276**, 42557–42564 (2001).

340. Tachu, B., Pillai, S., Lucius, R. & Pogonka, T. Essential Role of Chitinase in the Development of the Filarial Nematode Acanthocheilonema viteae . **76**, 221–228 (2008).

341. Wu, Y., Preston, G. & Bianco, A. E. Chitinase is stored and secreted from the inner body of microfilariae and has a role in exsheathment in the parasitic nematode Brugia malayi. *Mol. Biochem. Parasitol.* **161**, 55–62 (2008).

342. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).

343. Hecker, H. Structure and function of midgut epithelial cells in culicidae mosquitoes (insecta, diptera). *Cell Tissue Res.* **184**, 321–341 (1977).

344. Perzov, N., Padler-karavani, V., Nelson, H. & Nelson, N. Features of V-ATPases that distinguish them from F-ATPases. *FEBS Lett.* **504**, 223–228 (2001).

345. Johnson, L. S., Dunn, K. W., Pytowski, B. & Mcgraw, T. E. Endosome Acidification and Receptor Trafficking : Bafilomycin A1 Slows Receptor Externalization by a Mechanism Involving the Receptor ' s Internalization Motif. *Mol. Biol. Cell* **4**, 1251–1266 (1993).

346. Erickson, S. M. *et al.* Mosquito-Parasite Interactions Can Shape Filariasis Transmission Dynamics and Impact Elimination Programs. *PLoS Negl. Trop. Dis.* **7**, 1–7 (2013).

347. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag, 2016).

348. Ghedin, E. *et al.* Brugia malayi gene expression in response to the targeting of the Wolbachia endosymbiont by tetracycline treatment. *PLoS Negl. Trop. Dis.* **3**, (2009).

349. Liu, C., Kelen, P. Vander, Ghedin, E., Lustigman, S. & Unnasch, T. R. Analysis of transcriptional regulation of tetracycline responsive genes in Brugia malayi. *Mol. Biochem. Parasitol.* **180**, 106–111 (2011).

350. McGraw, E. A. & O'Neill, S. L. Wolbachia pipientis: Intracellular infection and pathogenesis in Drosophila. *Curr. Opin. Microbiol.* **7**, 67–70 (2004).

351. Comandatore, F. *et al.* Supergroup C Wolbachia, mutualist symbionts of filarial nematodes, have a distinct genome structure. *Open Biol.* **5**, 150099- (2015).

352. Sharma, V. *et al.* A Pilot Study of Bacterial Genes with Disrupted ORFs Reveals a Surprising Profusion of Protein Sequence Recoding Mediated by Ribosomal Frameshifting and Transcriptional Realignment. *Mol. Biol. Evol.* **28**, 3195–3211 (2011).

353. Abdel-Haleem, A. M. *et al.* The Emerging Facets of Non-Cancerous Warburg Effect. *Front. Endocrinol. (Lausanne).* **8**, (2017).

354. Shi, L. *et al.* Infection with Mycobacterium tuberculosis induces the Warburg effect

in mouse lungs. *Sci. Rep.* **5**, 1–13 (2015).

355. Degauque, N., Brosseau, C. & Brouard, S. Regulation of the Immune Response by the Inflammatory Metabolic Microenvironment in the Context of Allotransplantation. *Front. Immunol.* **9**, (2018).

356. Walker, T. *et al.* The wMel Wolbachia strain blocks dengue and invades caged Aedes aegypti populations. *Nature* **476**, 450–453 (2011).